



Universidade Federal de Santa Catarina  
Programa de Pós-Graduação em  
Engenharia de Produção

**UM MODELO PARA GERENCIAMENTO, AVALIAÇÃO E  
PLANEJAMENTO DA ARRECADAÇÃO DE  
TRIBUTOS ESTADUAIS**

Eugênio Rubens Cardoso Braz

Tese apresentada ao  
Programa de Pós-Graduação em  
Engenharia de Produção da  
Universidade Federal de Santa Catarina  
como requisito parcial para obtenção  
do título de Doutor em  
Engenharia de Produção

Florianópolis

2001

Eugênio Rubens Cardoso Braz

**UM MODELO PARA GERENCIAMENTO, AVALIAÇÃO E  
PLANEJAMENTO DA ARRECADAÇÃO DE  
TRIBUTOS ESTADUAIS**

Esta dissertação foi julgada e aprovada para a obtenção  
do título de Doutor em Engenharia de Produção  
no Programa de Pós-graduação em Engenharia de Produção  
da Universidade Federal de Santa Catarina

Florianópolis, 07 de abril de 2001

---

Prof. Ricardo Miranda Barcia , Ph.D.  
Coordenador do Curso

**BANCA EXAMINADORA**

---

Prof. Alejandro Rodriguez Martins, Dr.  
**Orientador**

---

Prof., Aran Bey Tcholakian, Dr.

---

Prof. Roberto C.S. Pacheco, Dr.

---

Prof. Malcon A. Tafner, Dr.

---

Prof. Marco A. Barbosa Cândido, Dr.

Ao meu pai, Pedro Alcântara Braz, *in memoria*

*Agradecimentos*

À Universidade Federal de Santa Catarina,  
À Coordenação de Aperfeiçoamento  
de Pessoal de Nível Superior CAPES,  
Ao Prof. Ricardo Miranda Barcia, Ph.D, por despertar o meu interesse,  
Ao orientador, Prof. Alejandro Rodriguez Martins, Dr.,  
Aos colegas que trabalharam comigo nos projetos  
e a todos os outros que direta ou indiretamente  
contribuíram para a realização  
deste trabalho.

## Resumo

BRAZ, Eugênio Rubens Cardoso. **Um modelo para gerenciamento, avaliação e planejamento da arrecadação de tributos estaduais**. Florianópolis, 2001. 156f.

Tese (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, UFSC, 2001.

Este trabalho analisa o problema da tributação no Brasil, e apresenta um modelo que permite o gerenciamento, avaliação e planejamento da arrecadação dos tributos de competência estadual, dando ênfase ao ICMS – Imposto sobre operações relativas à circulação de mercadorias e sobre prestações de serviços de transporte interestadual e intermunicipal e de comunicação, por ser este o tributo mais significativo para os cofres do governo.

Inicialmente, são apresentados os principais problemas enfrentados pela administração fazendária dos estados brasileiros, problemas estes relacionados por área (organização / gestão, tecnologia da informação, etc).

Segue-se com uma justificativa para a tributação e com uma revisão conceitual sobre as tecnologias de datawarehouse e datamining, usadas no modelo objeto desta tese.

A partir daí é descrito o modelo propriamente dito, que propõe e mostra como combater a sonegação fiscal e aumentar a receita estadual sem elevação da carga tributária.

Finalmente são apresentados resultados reais obtidos com um sistema aqui chamado de Inteligência Fiscal, desenvolvido à luz do modelo proposto, e aplicado a uma unidade da Federação.

## Abstract

BRAZ, Eugênio Rubens Cardoso. **A model for management, evaluation and planning of the collection of state tributes.** Florianópolis, 2001. 156 f.

Thesis (Doctorate in Production Engineering) – Program of Masters Degree in Production Engineering, UFSC, 2001.

This work deals with the taxation problem in Brazil, and presents a model that allows the management, evaluation and planning of the collection of tributes of state competence, giving emphasis to ICMS – Tax about relative operations to the circulation of goods and on installments of services of transport inter state and inter municipal and of communication, for being it the most significant tribute for the government's coffers. Initially, the main problems faced by the tributary administration of the Brazilian states are presented, problems these related by area (organization / administration, technology of the information, etc).

It is followed with a vindicative for the taxation and with a conceptual revision on the datawarehouse and datamining technologies, used in the model object of this thesis.

Starting from there, the model is described properly said, that proposes and shows as to combat the fiscal defraudment and to increase the state revenue without elevation of the tributary load.

Finally, real results obtained here with a system called of Fiscal Intelligence, developed based in the proposed model and applied to an unit of the Federation, are presented.

# ÍNDICE GERAL

|       |   |    |
|-------|---|----|
| 1     | INTRODUÇÃO .....  | 1  |
| 1.1   | Objetivo Geral .....  | 2  |
| 1.2   | Objetivos Específicos .....                                 | 2  |
| 1.3   | Justificativa e Importância do Trabalho .....               | 3  |
| 1.4   | Estrutura do Trabalho.....                                  | 4  |
| 2     | TRIBUTAÇÃO.....   | 6  |
| 2.1   | Conceitos fundamentais .....                                | 6  |
| 2.2   | Complexidade .....  | 10 |
| 2.3   | Logística do processo de arrecadação.....                   | 12 |
| 2.4   | Principais problemas .....                                  | 14 |
| 2.5   | Sistemas de controle tributário existentes .....            | 15 |
| 2.6   | Conclusão .....   | 19 |
| 3     | DATAWAREHOUSE E DATAMINING.....                             | 21 |
| 3.1   | Introdução.....   | 21 |
| 3.2   | Os princípios do Datawarehouse .....                        | 22 |
| 3.2.1 | Não Volatilidade .....                                      | 24 |
| 3.2.2 | Orientação por Assunto .....                                | 25 |
| 3.2.3 | Capacidade de Integração .....                              | 25 |
| 3.2.4 | Sensível ao Tempo.....                                      | 25 |
| 3.2.5 | Arquitetura Básica .....                                    | 26 |
| 3.2.6 | Arquitetura Centralizada.....                               | 27 |
| 3.2.7 | Arquitetura Distribuída .....                               | 28 |
| 3.2.8 | Principais aspectos a serem considerados.....               | 29 |
| 3.3   | Datamining .....  | 34 |
| 3.3.1 | Conceito.....   | 34 |
| 3.3.2 | Datamining e Datawarehouse .....                            | 39 |
| 3.3.3 | Classificação dos Métodos .....                             | 40 |
| 3.3.4 | Técnicas Utilizadas .....                                   | 41 |
| 3.3.5 | Métodos para Melhoria de Desempenho .....                   | 51 |
| 3.4   | Aplicações de datawarehouse e datamining .....              | 52 |
| 3.5   | Sistemas existentes .....                                   | 55 |
| 3.6   | Conclusão .....   | 58 |
| 4     | O MODELO PROPOSTO .....                                     | 60 |
| 4.1   | Introdução.....   | 60 |
| 4.2   | Etapas do Modelo .....                                      | 61 |
| 4.2.1 | Planejamento estratégico da necessidade de informações..... | 61 |
| 4.2.2 | DataWarehouse .....   | 76 |
| 4.2.3 | Datamining .....  | 93 |

|       |  |     |
|-------|--|-----|
| 5     | RESULTADOS OBTIDOS COM O MODELO.....   | 114 |
| 5.1   | Relativos ao Datawarehouse .....       | 114 |
| 5.1.1 | Contribuintes.....                     | 114 |
| 5.1.2 | Mercadorias .....                      | 116 |
| 5.1.3 | ICMS Declarado .....                   | 117 |
| 5.1.4 | ICMS Arrecadado .....                  | 117 |
| 5.1.5 | Outros Dados .....                     | 118 |
| 5.1.6 | Migração.....                          | 118 |
| 5.2   | Relativos ao Datamining .....          | 118 |
| 5.2.1 | Aspectos Contextuais.....              | 119 |
| 5.2.2 | Análise dos Tributos por Período ..... | 132 |
| 5.2.3 | Contribuintes a Serem Visitados .....  | 138 |
| 5.3   | Conclusão .....                        | 140 |
| 6     | CONCLUSÃO E RECOMENDAÇÕES .....        | 142 |
| 7     | REFERÊNCIAS BIBLIOGRÁFICAS.....        | 146 |



# 1 INTRODUÇÃO

A administração fazendária dos estados brasileiros, para fazer face às suas necessidades de investimento, vem estudando alternativas para aumentar a arrecadação de tributos. Nesta linha, esforços estão sendo concentrados para coibir a evasão de receita e divulgar junto a comunidade empresarial a função social do imposto, sua relevância no contexto administrativo, seus reflexos na construção do Estado e na qualidade de vida de seus habitantes.

As diferentes **SEF's**<sup>1</sup>, órgãos responsáveis pela administração fazendária estadual, encontram-se hoje em diferentes estágios nesta busca de alternativas de melhoria de suas gestões administrativo-tributárias. No entanto, os principais problemas enfrentados por elas, variando em intensidade, dizem respeito às mesmas questões, que estão relacionadas abaixo, por tópico: (PNAFE, Tocantins)

| QUESTÃO                  | PROBLEMAS  |
|--------------------------|--|
| Organização e Gestão     | Planejamento estratégico.                            |
|                          | Plano de cargos e salários.                          |
|                          | Instalações físicas.                                 |
| Tecnologia da informação | Plano diretor de informática.                        |
|                          | Tecnologia utilizada.                                |
|                          | Treinamento.   |
|                          | Comunicação de dados.                                |
|                          | Integração com outras organizações governamentais do |

---

<sup>1</sup> SEF – Secretaria Estadual da Fazenda.

|                            |  |
|----------------------------|--|
| Cadastro de contribuintes  | mesmo estado.<br><br>Integração com outras unidades da federação.                                      |
| Mercadorias                | Integração com outras unidades da federação.   |
| Fiscalização               | Planejamento da ação fiscal.<br><br>Programa de inteligência fiscal.                                   |
| Arrecadação                | Programa de previsão, acompanhamento e análise da receita.<br><br>Recepção eletrônica das declarações. |
| Contencioso Administrativo | Processo de julgamento.<br><br>Treinamento dos julgadores.   |

Estes problemas serão examinados no capítulo 2, que trata da tributação propriamente dita.

## 1.1 Objetivo Geral

O objetivo geral deste trabalho é idealizar e desenvolver um instrumento sobre gerência e inteligência fiscal, que contribua para o aprimoramento da gestão das finanças públicas feita pelas SEF's. Trata-se de um modelo de abrangência geral, que pode ser adaptado para atender necessidades específicas das SEF's de cada um dos estados brasileiros.

## 1.2 Objetivos Específicos

São considerados como objetivos específicos:

- Aumentar a receita do Estado, sem elevação de carga tributária.
- Possibilitar o acesso fácil aos dados que estão sendo utilizados.

- Gerar informações que subsidiem a tomada de decisões por parte das SEF's.
- Facilitar e tornar mais eficaz o trabalho dos profissionais da área fazendária (Fiscais, Delegados, Auditores, Gerentes e Secretário da Fazenda).
- Detectar os principais indícios de evasão fiscal.
- Descobrir e relacionar os contribuintes que possuem indícios de evasão fiscal.
- Dificultar a ação dos sonegadores (política preventiva).

### **1.3 Justificativa e Importância do Trabalho**

O assunto tema deste trabalho se enquadra numa ampla área referenciada hoje como modernização administrativa, alvo da atenção dos governos estaduais e federal brasileiro. No caso específico da administração tributária (este caso), devido a influência que exerce sobre a obtenção dos recursos financeiros necessários ao Estado, foi criado pelo governo um Programa Nacional de Apoio à Administração Fiscal para os Estados Brasileiros, com financiamento do Banco Interamericano de Desenvolvimento. Este fato por si só é suficiente para demonstrar o interesse despertado e a importância deste tema. Mas, além disto, está aí a sociedade a clamar por justiça social, que só irá se concretizando na medida em que cada cidadão for contribuindo de maneira justa, de acordo com os seus deveres, conforme especificado na legislação tributária em vigor. Os mecanismos desenvolvidos neste trabalho estão direcionados para este objetivo, que em última instância visa a melhoria da qualidade de vida dos cidadãos.

Para dar uma idéia do montante perdido com a sonegação de impostos, cita-se que, segundo estimativa de profissionais da SEF de Santa Catarina, para cada duas unidades monetárias arrecadadas no Estado, uma é sonegada, ou seja, um índice de evasão de 33,3%. Considerando a arrecadação média mensal (ano 2000) de R\$ 235.000.000,00 (duzentos e trinta e cinco milhões), o Estado deixa de arrecadar aproximadamente R\$

78.000.000,00 (setenta e oito milhões) todos os meses devido a evasão fiscal, montante este superior a três vezes a arrecadação do Estado do Tocantins.

Este mesmo raciocínio pode ser aplicado para as demais unidades federativas onde os Estados podem ser enquadrados em três grupos, em função das suas arrecadações de ICMS (PPGEP, 1998):

O primeiro, formado exclusivamente por São Paulo, que sozinho é responsável por 38,5% da arrecadação global do país..

O segundo, formado por:

- Minas Gerais, com .....9,7%
- Rio de Janeiro, com .....9,4%
- Rio Grande do Sul, com .....7,2%
- Paraná, com .....5,0%
- Bahia, com .....4,5%
- Santa Catarina, com .....3,4%

O terceiro, que arrecada 22,3% do total global no país, é formado pelos demais 20 estados.

## **1.4 Estrutura do Trabalho**

Para a elaboração da apresentação desta tese, são utilizados sete capítulos e um anexo:

No Capítulo 1, (este capítulo), apresenta-se a visão geral do trabalho e as considerações iniciais que delinearão o contexto dos problemas existentes; são também apresentados os objetivos, a justificativa e a estrutura utilizada para apresentação escrita do trabalho.

No Capítulo 2, são apresentados os conceitos fundamentais ao entendimento do contexto onde se vai trabalhar, juntamente com os principais problemas enfrentados pela administração fazendária. É dada ênfase ao tributo ICMS, e apresenta-se a visão

logística do seu processo de arrecadação. São também apresentados em anexo os aspectos da Constituição da República Federativa do Brasil que estabelecem as diretrizes básicas do Sistema Tributário Nacional.

No Capítulo 3, são apresentadas considerações conceituais sobre as tecnologias de *DataWarehouse* e *DataMining*. O interesse principal é fazer uma revisão bibliográfica sobre estes assuntos.

O Capítulo 4 é dedicado à apresentação do modelo objeto deste trabalho.

O Capítulo 5 apresenta os resultados reais obtidos com um software concebido e implementado à luz do modelo proposto, aplicado a uma unidade da Federação Brasileira.

O Capítulo 6 fica encarregado da conclusão e das recomendações para futuros trabalhos.

No Capítulo 7 relaciona-se a bibliografia que foi pesquisada para a realização do trabalho.

## 2 TRIBUTAÇÃO

Neste capítulo são citados os aspectos da Constituição da República Federativa do Brasil que estabelecem as diretrizes básicas do Sistema Tributário Nacional. São também apresentados os conceitos fundamentais ao entendimento do contexto onde se vai atuar e os principais problemas enfrentados pela administração fazendária.

É dada ênfase ao tributo ICMS, e apresenta-se a visão logística do seu processo de arrecadação.

A fundamentação necessária para o processo brasileiro de tributação, está contida na Constituição da República Federativa do Brasil de 5 de outubro de 1988, onde poderão ser consultadas informações adicionais às que serão apresentadas aqui, que se restringirão àquelas estritamente indispensáveis para a caracterização do contexto que será utilizado neste trabalho.

Tendo esta idéia em mente, vai-se utilizar o **Título VI – Da Tributação e do Orçamento** que se inicia na página 89 da referida Constituição, e que está aqui apresentado no Anexo I.

### 2.1 Conceitos fundamentais

A palavra **tributação**, e alguns outros termos que são utilizados no Anexo I, precisam ter os seus significados bem entendidos. Com este objetivo, será apresentada agora uma visão geral da tributação, bem como uma síntese dos principais conceitos diretamente relacionados a ela (MARQUES, 1997).

Apesar de cada ser humano dedicar-se às suas próprias atividades, muitas delas pertencentes a esferas completamente diferentes, a verdade é que existe uma forte

interdependência entre eles, pois uns necessitam de outros para suas subsistências, e portanto precisam manter um relacionamento social.

A partir do momento desta constatação, surge a necessidade da elaboração de regras de conduta, que estabeleçam direitos e deveres de cada cidadão.

O homem cria então o Estado, e dentro dele uma Administração Pública para cuidar dos seus interesses comuns.

Obviamente esta Administração Pública tem um custo, não apenas devido ao pessoal envolvido na sua execução, como também devido à realização de obras e prestação de serviços para a comunidade.

Este custo é arcado pela própria sociedade, isto é, todos os beneficiários dão a sua contribuição, e esta contribuição é chamada de **tributo**.

Para disciplinar o poder (e o seu limite) do Estado cobrar tal contribuição das pessoas, e o dever (e seus limites) dessas pessoas pagarem, foi criado um ramo do Direito chamado Direito Tributário – “tudo para que ninguém pague mais do que deve, nem o Estado cobre mais do que pode”. (MARQUES, 1997)

Seguem algumas conceituações de Direito Tributário, feitas pelos seguintes especialistas:

- ❑ Ruy Barbosa Nogueira: “Direito Tributário é a disciplina da relação entre o Fisco e o Contribuinte, resultante da imposição, arrecadação e fiscalização dos impostos, taxas e contribuições”.
- ❑ Bernardo Ribeiro de Moraes: “Direito Tributário é o conjunto sistemático de princípios e normas jurídicas disciplinadoras do poder fiscal do Estado, nas relações com as pessoas à ele sujeitas”.

- ❑ Geraldo Ataliba: “Direito Tributário é o sub ramo do Direito Público que fixa os princípios e normas que regem as relações entre o Estado e os particulares, no que toca a atividade financeira daquele, tendo em vista a arrecadação de tributos”.
- ❑ Rubens Gomes de Souza (pai da codificação do Direito Tributário): “Direito Tributário é o ramo do Direito Público que rege as relações jurídicas entre o Estado e os particulares, decorrentes de atividades financeiras do Estado no que se refere à obtenção de receitas que correspondam ao conceito de Tributos”.

Neste ponto vale o destaque para alguns conceitos:

➤ **Tributos:**

São as receitas derivadas que o Estado recolhe do patrimônio dos indivíduos, baseado no seu poder, mas disciplinado por normas do Direito Público, que constituem o Direito Tributário. Geraldo Ataliba diz que “Juridicamente se define tributo como obrigação jurídica pecuniária, *ex lege*, que não se constitui em sanção de ato ilícito, cujo sujeito ativo é, em princípio, uma pessoa pública, e cujo sujeito passivo é alguém nessa situação posto pela vontade da lei” (MARQUES, 1997).

➤ **Imposto:**

É o tributo, cuja obrigação tem por fato gerador uma situação independente de qualquer atividade estatal específica, relativa ao contribuinte. Geraldo Ataliba diz: “Define-se o imposto como um tributo não vinculado, ou seja, tributo cuja hipótese de incidência consiste num fato qualquer que não se constitua numa atuação estatal. O critério seguro para se reconhecer o imposto é o critério da exclusão: se, diante de uma exação, o intérprete verifica que não se trata de tributo vinculado, então pode afirmar seu caráter de imposto. Todo tributo não vinculado é imposto”.



➤ **Taxa:**

É o tributo arrecadado em razão do exercício do poder de polícia, ou pela utilização efetiva ou potencial de serviços públicos específicos e divisíveis, prestados ao contribuinte ou postos à sua disposição, conforme define o art. 18, inc. II do Código Tributário Nacional.

Geraldo Ataliba diz: “Taxa é o tributo vinculado, cuja hipótese de incidência consiste numa atuação estatal direta e imediatamente referida ao obrigado”. Diferentemente do imposto, a taxa pressupõe uma ação do Poder Público: a cobrança de uma taxa é vinculada ou ao exercício do poder de polícia, ou ao custeio de serviços públicos específicos e divisíveis, que o Poder Público presta ao contribuinte, ou coloca à sua disposição potencial.

➤ **Contribuição de Melhoria:**

É o tributo arrecadado dos proprietários de imóveis valorizados por obras públicas, que terá como limite total a despesa realizada, e como limite individual o acréscimo do valor que da obra resultar para cada imóvel beneficiado. . Geraldo Ataliba diz: “É a contribuição de melhoria, instrumento de efetivação da atribuição da valorização imobiliária causada por obra pública, à pessoa que a empreendeu. Juridicamente se distingue do imposto e da taxa porque sua hipótese de incidência não é a obra (como na taxa), nem o mero enriquecimento do contribuinte (como no imposto), mas a diferença do valor de uma propriedade antes e depois da obra. Para que se configure o fato imponible, não basta que haja obra (taxa), nem basta que haja incremento patrimonial (imposto). É preciso haver direta relação de causa e efeito entre a obra e a valorização”.

Nota-se que o fato gerador é quem determina a natureza do tributo: se é imposto, se é taxa ou se é contribuição de melhoria.

Pode-se dizer ainda, que, para a determinação da natureza do tributo, o nome que lhe é atribuído não terá a menor importância; veja o exemplo da Taxa de Pavimentação, que tem natureza jurídica de contribuição de melhoria, ou do Selo Pedágio, que tem natureza jurídica de imposto (MARQUES, 1997).

## **2.2 Complexidade**

Não obstante existam fontes de informações que direcionam os procedimentos a serem adotados no âmbito do ICMS, a complexidade do problema de gerenciamento e controle da tributação é de tal forma alta, que, tanto facilita as ações de sonegação por parte de contribuintes, como dificulta as ações de fiscalização por parte da Secretaria da Fazenda, órgão responsável pela administração tributária.

Desta forma, além de um sólido conhecimento sobre os conceitos apresentados anteriormente, os profissionais da Fazenda precisam, para fazer frente ao problema da tributação, estar bem familiarizados com os tópicos relacionados em seguida: (RICMS, 1997 - Regulamento do ICMS)

- Da Incidência do Imposto e do Fato Gerador
- Da Não-Incidência
- Das Isenções, Incentivos e Outros Benefícios Fiscais
- Da Base De Cálculo
- Das Alíquotas
- Da Compensação do Imposto
- Do Lançamento e do Recolhimento do Imposto
- Do Local da Operação e da Prestação

- Do Estabelecimento
- Da Sujeição Passiva
- Das Operações e Prestações Especiais
- Das Rotinas de Controle e Fiscalização de Mercadorias Objeto de Serviço Postal
- Das Operações de Importação e de Arrematação de Mercadorias Procedentes do Exterior
- Das Operações Realizadas por Concessionários, Revendedores, Agências e Oficinas Autorizadas de Veículos, Tratores, Máquinas, Eletrodomésticos e Outros Bens
- Do Cancelamento de Benefícios Fiscais e da Cassação de Regime Especial para Pagamento, Emissão de Documentos Fiscais ou Escrituração de Livros
- Das Infrações e das Penalidades
- Da Fiscalização
- Do Documentário e da Escrita Fiscal
- Da Emissão dos Documentos Fiscais e Escrituração de Livros Fiscais por Contribuinte Usuário de Sistema Eletrônico de Processamento de Dados
- Impressão e Emissão Simultânea de Documentos Fiscais
- Do Cadastro de Contribuinte do Estado
- Máquinas Registradoras
- Do Uso de Terminal Ponto de Venda – PDV
- Do Equipamento Emissor de Cupom Fiscal – ECF
- Dos Regimes Especiais de Tributação, Escrituração de Livros e Emissão de Documentos Fiscais
- Das Operações com o Fim Específico de Exportação
- Da Substituição Tributária

- Das Operações Sujeitas a Antecipação Tributária
- Da Certidão Negativa de Débitos Tributários
- Do Código Fiscal de Operações e Prestações, do Código de Atividades Econômicas e do Código de Situação Tributária
- Das Disposições Gerais, Finais e Transitórias

Cada um destes tópicos apresentados acima, requer exames e análises criteriosos para evitar ambigüidade de interpretações.

### 2.3 Logística do processo de arrecadação

Conforme foi visto no início deste capítulo, existem diretrizes básicas que norteiam o processo de tributação, diretrizes estas expressas na Constituição da República Federativa do Brasil e em documentos mais detalhados, produzidos a partir dela, como por exemplo, o **Sistema Tributário Nacional**, que estabelece os princípios gerais (da estrita legalidade, da anterioridade, da irretroatividade da lei tributária, da estrita igualdade e da uniformidade geográfica), as competências tributárias, as limitações ao poder de tributar, etc, e o **Código Tributário Nacional**, lei complementar que traça as normas gerais do direito tributário, que são:

- Definição de tributos e de suas espécies, bem como, em relação aos impostos discriminados na Constituição, a dos respectivos fatos geradores, base de cálculo e contribuintes.
- Obrigação, lançamento, crédito, prescrição e decadência tributários.
- Adequado tratamento tributário ao ato cooperativo praticado pelas sociedades cooperativas.

No caso específico do ICMS, objeto deste trabalho, ainda existe o “Regulamento do Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de

Serviços de Transporte Interestadual e Intermunicipal e de Comunicação” (RICMS), a que se refere o Decreto nº 13.640, de 13 de Novembro de 1997.

Uma etapa fundamental para o processo de tributação, é a logística para a coleta dos documentos fiscais (PPGEP, 1998), que serão utilizados como dados de entrada. A Figura 1, apresentada em seguida, ajuda o seu entendimento.

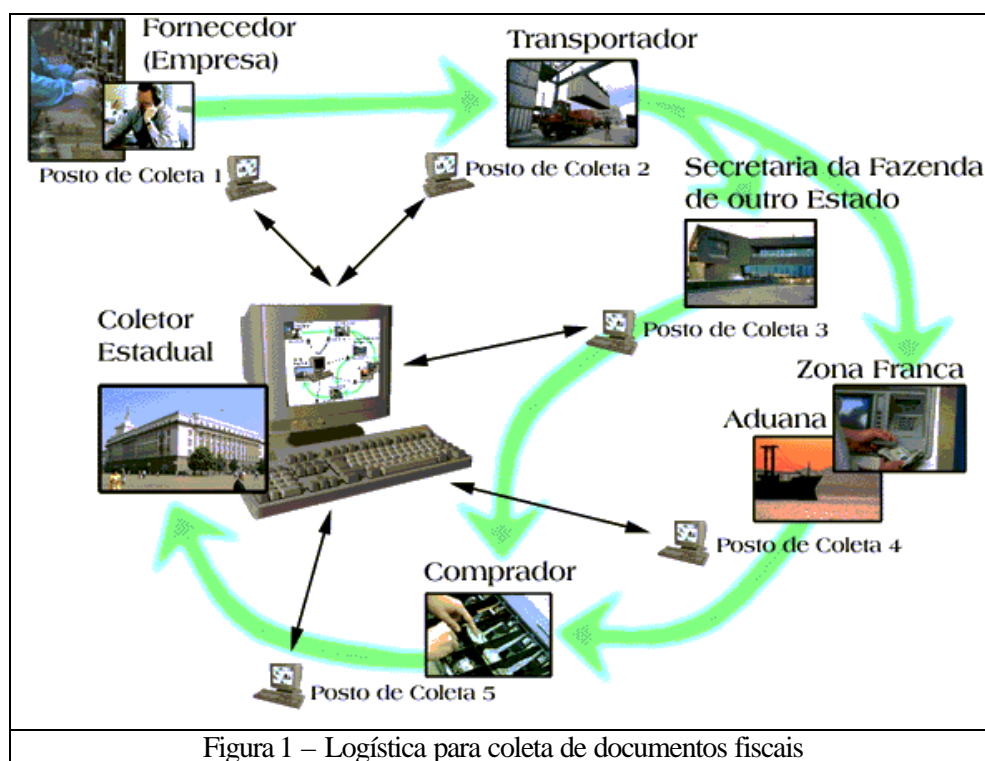


Figura 1 – Logística para coleta de documentos fiscais

Como se pode ver na Figura 1 (PPGEP, 1998), existem vários pontos de coleta de documentos fiscais, genericamente chamados de postos fiscais, distribuídos ao longo de um percurso, que se inicia sempre num fornecedor de algum produto ou serviço, e termina invariavelmente num comprador.

Existe uma diversidade de processos de recepção dos dados nos postos fiscais, de acordo com o grau de informatização da empresa fornecedora, que variam desde documentos escritos em papel até aqueles já em meio legível por computador.

Todos estes dados são enviados para um coletor estadual, utilizando também para isto diferentes meios: formulários em papel, diskette, internet, intranet, etc.

## 2.4 Principais problemas

Os principais problemas existentes no âmbito da administração fazendária, variando em intensidade em função das características particulares de cada Estado, se enquadram sempre nas áreas relacionadas abaixo (PNAFE, Tocantins):

- **Organização e Gestão:** Falta de um planejamento estratégico integrado, Política salarial que não incentiva a melhoria de desempenho, Instalações físicas não muito adequadas às condições de trabalho, Ausência ou insuficiência de treinamentos voltados para a capacitação gerencial e operacional, Inexistência de programas de qualidade e de melhoria contínua.
- **Tecnologia da Informação:** Falta de um Plano Diretor de Informática, Utilização de tecnologia já superada, Deficiência nos programas de capacitação profissional, tanto da área de informática, quanto da área dos usuários, Equipamentos inadequados e Comunicação de dados pouco eficiente.
- **Cadastro dos Contribuintes:** Falta de integração com outras organizações governamentais do mesmo Estado, como por exemplo, a Junta Comercial, a Companhia de Energia Elétrica e a Companhia Telefônica, Falta de mecanismos de depuração e atualização sistemática do cadastro, Controle deficiente da autorização da impressão de documentos fiscais, Falta de integração com outras unidades da Federação, com o Renavam, etc.
- **Mercadorias:** Falta de integração com as demais unidades da Federação, o que dificulta sobremaneira a sistemática de verificação da confiabilidade dos dados,

Deficiência nos processos de aquisição dos dados nos pontos remotos, incluindo-se aí a logística, a crítica e a transmissão.

- **Fiscalização:** Falta de um programa de Inteligência Fiscal, que ajude no planejamento e gerenciamento da ação fiscal, Falta da informatização de normas e procedimentos fiscais, Falta de estrutura e equipamentos de apoio nas unidades de fiscalização de mercadorias em trânsito e Falta de um programa contínuo de capacitação.
- **Arrecadação:** Falta de um programa de previsão, acompanhamento e análise da receita, Instalações inadequadas para o funcionamento das coletorias, Elevado número de documentos fiscais para o produtor rural, Conciliação da arrecadação com o Guia Nacional de Recolhimento difícil e demorada e Falta de um programa de auditoria na rede bancária.
- **Contencioso Administrativo:** Falta de treinamento dos julgadores, Lentidão do processo de julgamento, Excessiva demora na restituição do indébito tributário e Deficiência na divulgação das decisões definitivas.

## 2.5 Sistemas de controle tributário existentes

### No Brasil

Realizou-se uma extensiva análise dos sistemas existentes no mercado nacional e internacional que se propõem a auxiliar a administração tributária e a identificar focos de sonegação de impostos no âmbito governamental. A Tabela 1 e a Tabela 2 apresentam exemplos de sistemas desenvolvidos por órgãos de informática da administração pública e por empresas privadas, respectivamente.

Tabela 1: Sistemas desenvolvidos por órgãos de informática da administração pública

| Instituição  | Sistema  | Descrição  |
|--|--|--|
| SERPRO -<br>Ministério da<br>Fazenda<br><a href="http://www.serp.ro.gov.br/">http://www.serp.ro.gov.br/</a>                  | <ol style="list-style-type: none"> <li>1. ANGELA. Análise Gerencial da Arrecadação</li> <li>2. AUDITA. Sistema de Auditoria Fiscal</li> <li>3. SIPRAR. Sistema de Previsão e Análise da Arrecadação</li> <li>4. SIGA. Sistema de Informações Geradoras de Ação Fiscal</li> </ol> | <ol style="list-style-type: none"> <li>1. Conjunto de arquivos contendo dados analíticos e agregados no período de 24 meses.</li> <li>2. Sistema de apoio ao Auditor Fiscal nas investigações contábeis/fiscais de um determinado contribuinte</li> <li>3. Conjunto de arquivos contendo dados de previsão da Arrecadação Federal.</li> <li>4. Base de informações sobre contribuintes - Pessoa Física, relativa ao seu potencial econômico e imposto suplementar a pagar, integrando diversas bases da Receita Federal e de outras entidades</li> </ol> |
| CELEPAR<br><a href="http://www.pr.gov.br:80/celepar/celepar/catalogo/">http://www.pr.gov.br:80/celepar/celepar/catalogo/</a> | FIS. Controle de Notas Fiscais (Âmbito estadual)   | Gerencia as notas fiscais emitidas pelos contribuintes. Está integrado com alguns sistemas, como: biblioteca de legislação tributária (BLT), conta corrente fiscal (CCF), cadastro de contribuintes do ICMS (CIF), dívida ativa fiscal (DAF), informação de arrecadação do ICMS (FIR), etc.  |
| CELEPAR<br><a href="http://www.pr.gov.br:80/celepar/celepar/catalogo/">http://www.pr.gov.br:80/celepar/celepar/catalogo/</a> | DCR. Controle da Receita (Âmbito estadual)   | Cadastra as guias de arrecadação referentes a taxas, multas e licenciamentos, gerando controles para fins estatísticos e de repasse financeiro aos órgãos envolvidos.  |
| CIASC<br><a href="http://www.ciasc.gov.br/">http://www.ciasc.gov.br/</a>   | DIF. Declaração de Informações Econômico-Fiscais (Âmbito estadual)   | Apura os índices de participação dos municípios no produto da arrecadação de ICMS. O sistema contém as informações econômico-fiscais apresentadas anualmente por todas as empresas contribuintes do ICMS do estado. É integrado com os sistemas de Arrecadação e Cadastro de Contribuintes   |



|  |  |  |
|--|--|--|
| CIASC<br><a href="http://www.ciasc.gov.br/">http://www.ciasc.gov.br/</a>                           | AIE. Sistema de Autorização para Emissão de Documentos Fiscais (Âmbito estadual) | O sistema possui uma base de dados em que são armazenadas as informações das notas fiscais dos contribuintes do Estado. Controla e fiscaliza o uso de notas fiscais. |
| PRODERJ<br><a href="http://www.proderj.rj.gov.br/rar.htm">http://www.proderj.rj.gov.br/rar.htm</a> | RAR/RFA/RRA. Arrecadação Estadual (Âmbito estadual)                              | Valida a arrecadação de tributos estaduais e mantém uma base de dados com os tributos arrecadados. Possibilita a entrada de dados e consultas diversas.              |

Ressalta-se que os sistemas existentes na esfera estadual, tanto os desenvolvidos por empresas privadas quanto por órgãos de informática da administração pública, limitam-se a atividades de manutenção do sistema atual de gerenciamento tributário e/ou automatização de atividades do processo fiscal. Estes sistemas constituem, portanto, coletores de dados que facilitam determinadas etapas do processo fiscal, sem relacionar fatores de irregularidade. As bases de dados geradas, no entanto, estão distribuídas e, algumas vezes, desintegradas. Trata-se de um dos fatores de multiplicação de informações na administração pública. A multiplicidade de sistemas independentes é o maior fator de desagregação da informação em uma organização.

Tabela 2: Sistemas desenvolvidos por empresas privadas.

|   |   |  |
|---|---|--|
| Banfiscal Empresa Jornalística e Editora Tributária Eletrônica<br><a href="http://www.banfiscal.com.br/index.htm">http://www.banfiscal.com.br/index.htm</a> | Banco de Dados Jurídico (Âmbito nacional) | Permite a consulta “on-line” de legislação atualizada, com acesso instantâneo à informação.  |
| Pólis Informática<br><a href="http://www.vanet.com.br/polis/">http://www.vanet.com.br/polis/</a>  | Muni/Cerebrum Tributos (Âmbito municipal) | Sistema para a administração e controle da tributação, arrecadação e fiscalização. Subdividido nos seguintes módulos: cadastro de contribuintes, cadastro imobiliário, cadastro de atividades, módulo IPTU, módulo ISQN, módulo certidões e notificações, módulo |

|   |  |   |
|---|--|---|
|   |  | ITBI, módulo taxas e serviços, módulo de contribuição de melhorias.   |
| SMAR-ADP<br><a href="http://www.smarapd.com.br/">http://www.smarapd.com.br/</a> | Sistema de Tributação Municipal (Âmbito municipal) | Objetiva racionalizar o uso de recursos na administração municipal. Processa e emite documentos de arrecadação municipal. Realiza cadastramentos municipais, calcula impostos, controla a arrecadação. Emite livro de dívida ativa, emite relatórios diversos e mantém cadastro geral de CGC/CPF. |

### **No Exterior**

Referindo-se a outros países, precisa-se antes de mais nada, levar em consideração o estágio de desenvolvimento econômico dos mesmos, pois, de acordo com a teoria econômica tradicional, este é o fator determinante da tributação.

Os países em estágios iniciais de desenvolvimento, caracterizam-se via de regra, por reduzido mercado interno e grande dependência dos fluxos internacionais de bens e serviços. A principal fonte de recursos tributários situa-se no comércio exterior (OLIVEIRA, 1998).

Os países que experimentam um processo continuado de desenvolvimento, caracterizam-se pela adoção de políticas de substituição de importações, de incentivos à formação de um parque industrial e de ampliação do seu mercado interno. Esses países tendem a tributar mais intensamente a produção e a circulação, pois as transações internas passam a ser mais significativas do que as transações internacionais (OLIVEIRA, 1998).

Os países mais desenvolvidos, detentores de um parque industrial mais moderno e diversificado, além de manterem a incidência sobre o consumo, tendem a privilegiar a

tributação sobre a renda e o patrimônio, de modo a fazer com que cada indivíduo ou empresa recolha o tributo em função de sua capacidade econômica (OLIVEIRA, 1998).

As soluções apresentadas no mercado americano, por exemplo, não contemplam os procedimentos de análise utilizados neste trabalho e sim sistemas de apoio logístico às atividades de arrecadação. Muito embora a realidade dos Estados Unidos seja muito diferenciada da Brasileira, investigou-se também o mercado americano por sistemas que identifiquem focos de sonegação fiscal. A análise revelou apenas a existência de sistemas de automatização do cálculo e atualização de impostos, como os desenvolvidos pela Auto Administrator Int'l Inc. (<http://www.autoadmin.com/>), e The Logics Tax Billing System (<http://www.logics-software.com/index.htm>), para citar apenas alguns.

Entretanto, tanto a crescente demanda e popularidade de sistemas baseados em técnicas de *DataMining* como a preocupação dos governos estaduais e federal americanos com a evasão fiscal<sup>2</sup>, podem se configurar em fatores que estimulem o desenvolvimento de sistemas para inteligência fiscal naquele país. Porém, a diferença na logística de arrecadação, nos estágios de desenvolvimento e na legislação entre os sistemas fiscais brasileiro e americano, impedem a aplicabilidade de tais sistemas no Brasil.

## 2.6 Conclusão

A evasão fiscal é um problema que vem desafiando a administração tributária brasileira há muito tempo; com o seu crescente agravamento, e suas consequências nefastas para o governo e para a população, esforços estão sendo dispendidos em todos os âmbitos governamentais, visando coibir esta evasão de receita e divulgar junto a comunidade

---

<sup>2</sup> O departamento do tesouro americano, o *Internal Revenue Service*, já manifestou interesse no *Data Mining* para identificar padrões de irregularidades nas declarações de imposto de renda (M. J. A. Berry and G. Linoff, *Data Mining Techniques: for marketing, sales and customer support*, John Wiley & Sons, 1997).

empresarial a função social do imposto, sua relevância no contexto administrativo, seus reflexos na construção do Estado e na qualidade de vida de seus habitantes.

Neste capítulo, procurou-se conhecer o contexto onde está inserida a administração tributária, as leis e regulamentos que a rege, os principais problemas que a aflige e os recursos tecnológicos atualmente disponíveis.

### 3 DATAWAREHOUSE E DATAMINING

#### 3.1 Introdução

A capacidade que tinha o homem de produzir dados aumentou de uma maneira surpreendente na última década do século XX, e as perspectivas indicam que esta tendência de crescimento continuará sendo uma tônica dominante neste novo milênio.

Os recursos tecnológicos atualmente disponíveis facilitam enormemente o processo de coleta de dados, vide aí a Internet, e sugere o desenvolvimento de novas tecnologias capazes de tratar estes dados, transformá-los em informações úteis e extrair conhecimentos (*knowledge discovery*) a partir deles. (BRACHMAN e ANAND, 1996).

Milhares de bases de dados voltadas para administração pública, gerenciamento de negócios, aplicações científicas, engenharia e muitas outras, vem sendo usadas por organizações governamentais, empresas privadas, universidades e centros de pesquisas.

Grande parte destas organizações está vivenciando a crise dos dados, que vem sendo provocada pelo armazenamento de dados heterogêneos “*disparate data*” (BRACKETT, 1996) ao longo do tempo, crise esta que poderá culminar com a perda total do controle sobre os dados, com reflexos imediatos na competitividade de mercado (ARMSTRONG, 2000).

Pesquisadores de diferentes áreas (sistemas de base de dados, inteligência artificial, sistemas baseados em conhecimentos, *machine learning* (WEISS e KULIKOWSKI, 1991), estatística, aquisição de conhecimentos, base de dados espaciais, visualização de dados, etc.), vem considerando o fato de obter informações e extrair conhecimentos a partir de grandes massas de dados, como sendo um tópico chave de pesquisa, e devido

à sua importância, têm demonstrado interesse do assunto, que está universalmente referenciado como *Datamining* (CHEN et al. 1996).

Para facilitar o trabalho supracitado, torna-se fundamental uma criteriosa análise dos dados armazenados nas diversas bases de dados, visando reagrupá-los e reorganizá-los estrategicamente, de acordo com finalidade e critérios previamente estabelecidos. Surge então o *Datawarehouse*, que pode ser visto como uma fonte de dados voltada para o suporte à decisão de usuários finais, fonte esta derivada de diversos bancos de dados operacionais (SIMON, 1995), ou como um conjunto de dados baseado em assuntos, integrado, não-volátil e variável em relação ao tempo, de apoio às decisões gerenciais. (INMON, 1997).

Este capítulo tem por objetivo apresentar conceitualmente os dois tópicos mencionados, e dar uma visão geral da tecnologia onde eles estão inseridos.

### **3.2 Os princípios do Datawarehouse**

Existem sistemas que foram desenvolvidos para atender as necessidades de informação das organizações. Estes sistemas são conhecidos como *Decision Support Systems* – DSS e *Executive Information Systems* – EIS.

Satisfazer ambas as necessidades, operacionais e de informação de uma organização, é uma tarefa muito difícil. Bases de dados estruturadas para atender necessidades operacionais mostram-se inadequadas para analisar informações que dão suporte à tomada de decisões (SIMON, 1995).

Um *Datawarehouse*, contudo, variando o grau de sumarização dos dados operacionais, mantendo informações históricas e consolidando uma variedade de fontes de dados, pode resolver o problema de gerenciar dados para aplicações que produzam

informações executivas (BRACKETT, 1996). *Datawarehouse* é uma tecnologia emergente dos anos 90, e pode ser conceituada como:

"Uma fonte de dados logicamente integrados, voltada para aplicações de sistemas de suporte a decisão e de informações executivas, fonte esta derivada de diversas bases de dados operacionais" (SIMON, 1995).

Convém ressaltar que o *datawarehouse* não é uma base de dados, apesar de poder ser implementado usando-se um DBMS<sup>3</sup> ou um DDBMS<sup>4</sup>.

Como o objetivo de um *datawarehouse* é dar suporte ao processo de tomada de decisão, e não às operações ou transações da organização, alguns dos princípios utilizados na tecnologia de base de dados e de DBMS não são necessários aqui. Especificamente, as operações tradicionais de *update*, frequentes em bases de dados, não são utilizadas no *datawarehouse* (SIMON, 1995).

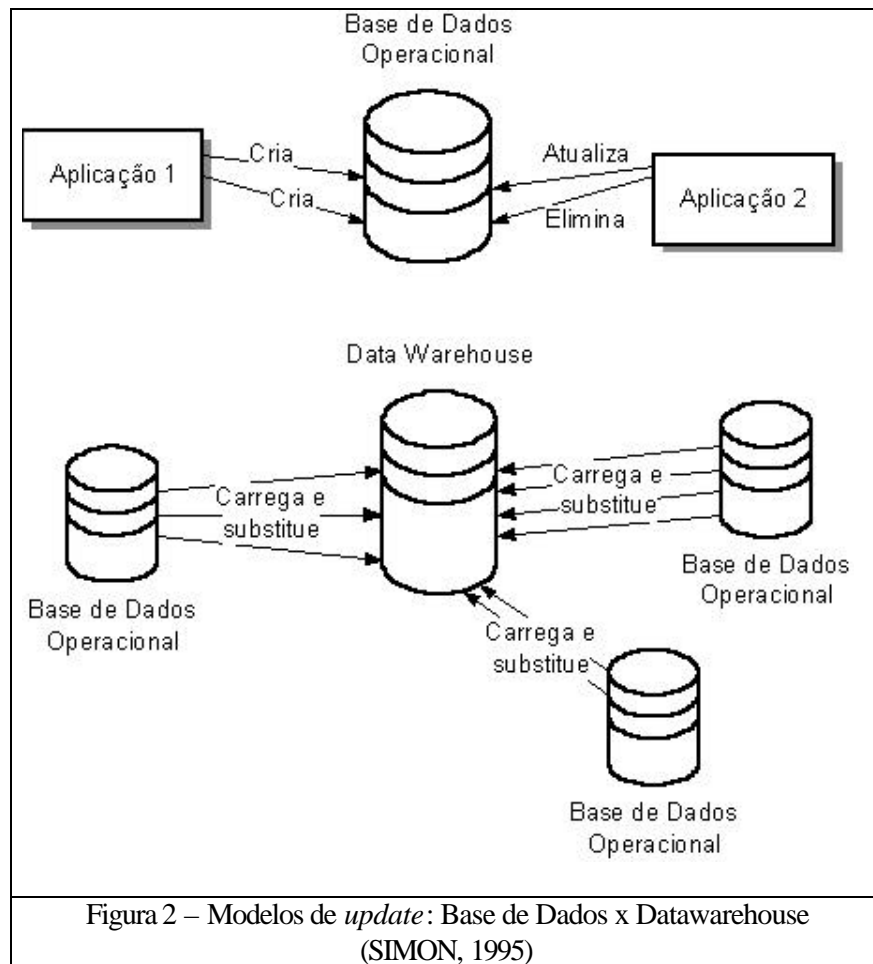
A Figura 2 mostra uma base de dados operacional típica, independente do modelo que ela utiliza (orientado a objeto (GIOVINAZZO, 2000), relacional ou hierárquico).

Diferentes ambientes têm diferentes níveis de volatilidade nos seus dados.

---

<sup>3</sup> Database management systems

<sup>4</sup> Distributed database management systems



### 3.2.1 Não Volatilidade

O *datawarehouse* é **não volátil** (INMON, 1997, SIMON, 1995), significando dizer, que ele não está sujeito às operações de atualização, tão freqüentes em base de dados operacionais; isto é um dos princípios que se deve ter em mente quando do seu projeto e construção.

Por exemplo, no modelo relacional ocorre com uma determinada freqüência, a inclusão e a eliminação de registros, a substituição de valores dos atributos e outras alterações que se tornem necessárias. Observando a Figura 2, é possível ver que estas operações não são utilizadas no *datawarehouse*, que, ao invés disto, recebe toda uma carga de



dados à intervalos regulares de tempo, obedecendo algumas regras de extração. Estes dados podem ser provenientes de uma base de dados centralizada, residente no mesmo ambiente do *datawarehouse*, como também podem envolver operações de extração e carga a partir de bases de dados distribuídas, residentes em ambientes diferentes.

O processo de carga usualmente envolve um sofisticado tratamento para eliminação de inconsistências de tipos de dados, tamanhos, significado dos atributos, codificação e outras propriedades intrínsecas dos dados que estão sendo recuperados (KIMBALL, REEVES, ROSS and THORNTHWAITE, 1998).

Após a carga dos dados, o *datawarehouse* fica pronto para atender solicitações de consultas feitas por sistemas de informações executivas e por sistemas de suporte à decisão.

### 3.2.2 Orientação por Assunto

Outro princípio básico, é que os *datawarehouses* são **orientados por assunto**, (INMON, 1997) e são organizados utilizando-se sub conjuntos particulares de dados, contidos em base de dados operacionais. São construídos extraindo-se dados de diferentes aplicações, que podem residir em diferentes plataformas, requerendo capacidade de integração, que é outro princípio de fundamental importância. Vide Figura 3 (SIMON, 1995).

### 3.2.3 Capacidade de Integração

A integração consiste em montar um esquema global e unívoco, a partir de múltiplas aplicações e fontes de dados diferentes, que utilizam critérios próprios e não uniformes entre si (INMON, 1997).

### 3.2.4 Sensível ao Tempo

Um outro princípio relevante é o *time variance*, pois o *datawarehouse* irá refletir sempre um momento no tempo; é como se fosse uma série de fotografias tiradas em épocas diferentes e consecutivas (INMON, 1997).

### 3.2.5 Arquitetura Básica

A Figura 3 mostra uma arquitetura básica de *datawarehouse*. Nela, dados contidos num ambiente operacional, composto por uma ou mais bases de dados, são extraídos de acordo com certas regras (FIRESTONE, 2000) e são carregados periodicamente no *datawarehouse*.

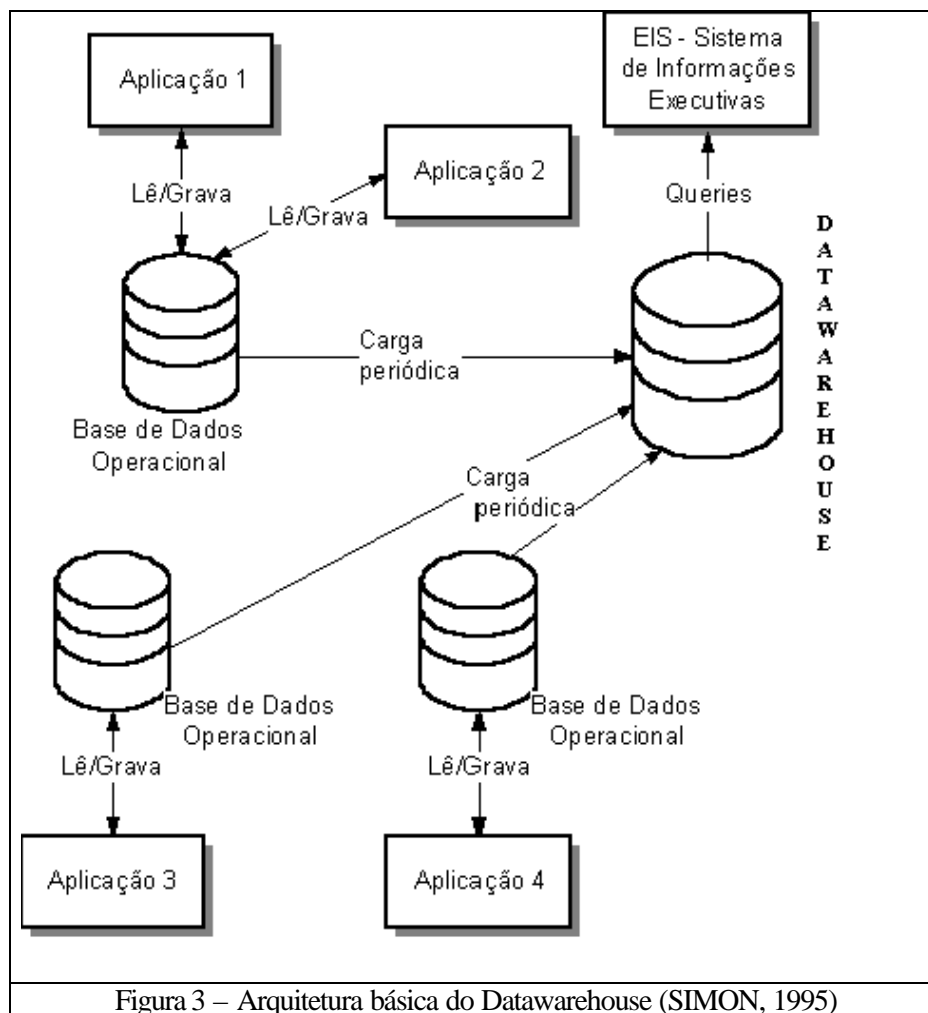
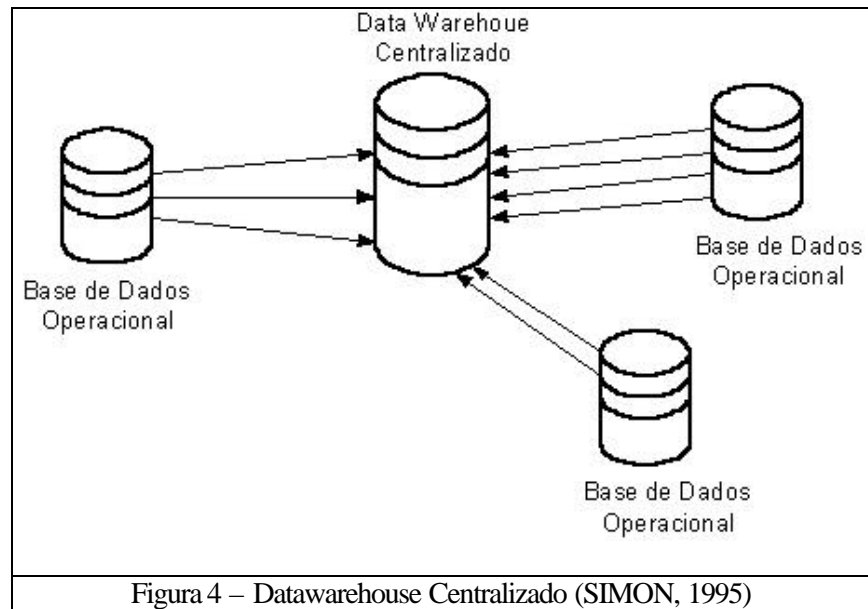


Figura 3 – Arquitetura básica do Datawarehouse (SIMON, 1995)

### 3.2.6 Arquitetura Centralizada

O *datawarehouse* pode ser centralizado, e esta arquitetura é apropriada para ambientes onde o *main frame* é também o foco dos dados da organização (Figura 4), ou pode ser distribuído.



A maioria das organizações constrói e mantém um único ambiente de *datawarehouse* centralizado (INMON, 1997) pelas seguintes razões:

Os dados contidos no *datawarehouse* são integrados através da empresa, e é somente na matriz que uma visão integrada é utilizada.

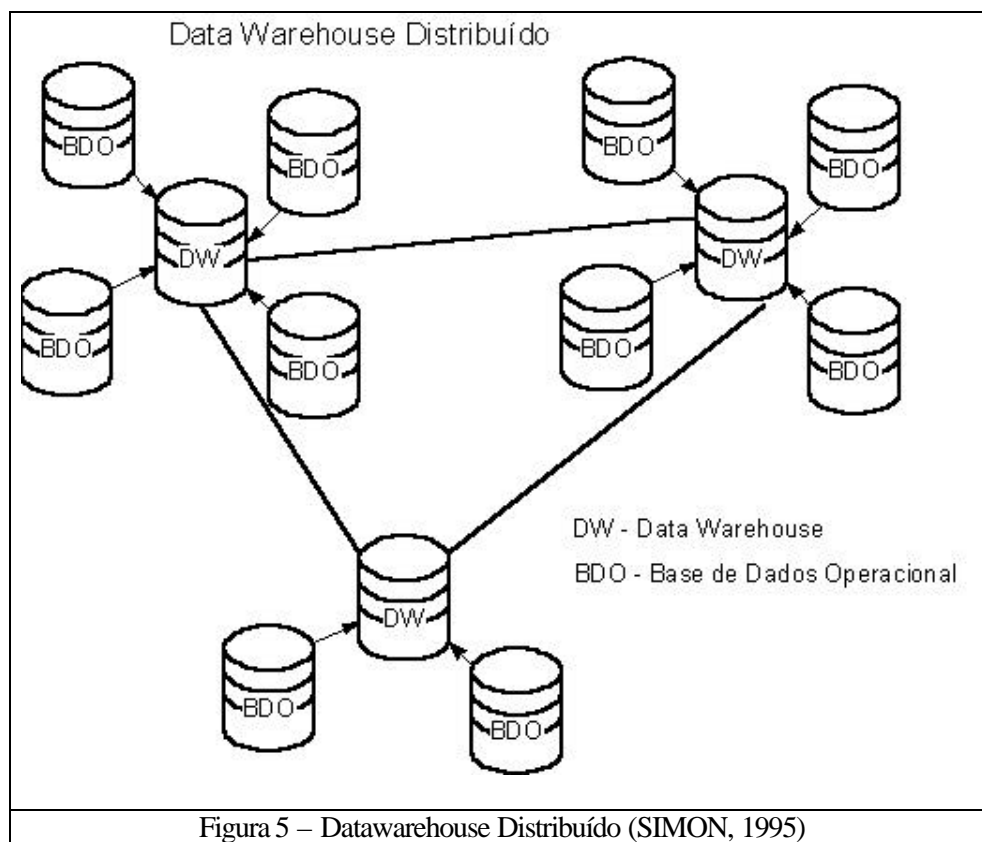
O volume de dados existentes na *datawarehouse* é tal que um único repositório de dados centralizado faz sentido.

Mesmo se os dados pudessem ser integrados, caso eles fossem espalhados por diversos *sites* locais, eles seriam de difícil acesso.

A política, a economia e a tecnologia favorecem amplamente a existência de um único *datawarehouse* centralizado.

### 3.2.7 Arquitetura Distribuída

Esta arquitetura é recomendada (FIRESTONE, 2000) quando as aplicações operacionais utilizarem recursos computacionais geograficamente distribuídos (Figura 5).



O *datawarehouse* distribuído faz sentido quando ocorre uma quantidade significativa de processamento nos diversos níveis locais, isto é, os *sites* locais têm autonomia de processamento e só ocasionalmente, para certos tipos de processamentos é que os dados são enviados para a central. (INMON, 1997)

### 3.2.8 Principais aspectos a serem considerados

Um dos aspectos considerados fundamentais para o *Datawarehouse* é a construção do que se chama de metadados (BRACKETT, 1996).

#### **Metadados**

São informações sobre os dados contidos no *Datawarehouse*, incluindo inclusive regras para utilização dos mesmos.

O gerenciamento dos metadados cresce em importância com o volume e a complexidade dos dados sob o seu controle. Bons metadados são decisivos para localizar e entender os dados necessários ao atendimento da demanda de informações da organização (BRACKETT, 1996).

As pessoas que desenvolvem os sistemas de informação e suas fontes de dados, conhecem os dados com os quais estão trabalhando. Quando estas pessoas mudam de trabalho ou se aposentam, este conhecimento é perdido. O que resta é um entendimento implícito sobre os dados utilizados pelos sistemas de informação.

A tendência contínua de perda de conhecimento sobre os dados com as mudanças das pessoas, é apenas parte do problema. A outra parte é que o conhecimento ainda existente fica tão heterogêneo quanto os dados reais, na medida em que vai sendo documentado em diferentes lugares, em diferentes formas e em diferentes graus de detalhe.

Toma-se difícil encontrar todos os metadados e integrá-los para obter um entendimento consistente sobre os dados reais.

As organizações precisam ter o melhor conhecimento possível sobre os seus dados, para que possam utilizá-los no atendimento às demandas de informações dos seus negócios (BRACKETT, 1996).

Para isto, é preciso que os metadados sejam vistos como parte do recurso integrado de dados, como os dados pessoais, financeiros, de clientes, etc.

Outro fator de sucesso ou fracasso para o *Datawarehouse* é a definição da granularidade dos seus dados (ZIMMER, 2001).

### **Granularidade**

É o nível de detalhe ou de síntese utilizado nos dados que serão armazenados. Quanto mais detalhe, menor é o nível de granularidade, e quanto menos detalhe, maior é o nível de granularidade.

A importância da granularidade, considerada a principal questão do projeto do *Datawarehouse* (INMON, 1997), consiste no fato dela afetar profundamente o volume de dados que irá para o *Datawarehouse*, e afetar também o tipo de consultas que poderão ser atendidas.

O volume de dados a ser armazenado é definido em função dos tipos de consultas que se desejam fazer.

Quando a granularidade é corretamente definida, se dá um grande passo na direção do sucesso do *Datawarehouse*, pois os demais aspectos do projeto ficam facilitados; no entanto, quando a granularidade é mal definida, ocorre exatamente o contrário (ZIMMER, 2001).

O equilíbrio na escolha dos níveis adequados de granularidade, é obtido entre o gerenciamento do volume de dados e o armazenamento de dados num nível tão alto de granularidade, que impossibilite a posterior consulta detalhada.

O ponto de partida para a definição do nível apropriado de granularidade é fazer uma estimativa bruta no número de registros que o *Datawarehouse* conterà.

Isto pode ser realizado através de um algoritmo muito simples, que consiste em identificar todas as entidades que serão utilizadas e, para cada uma delas, o tamanho do registro.

Em seguida estima-se a quantidade mínima e máxima de registros que serão utilizados no horizonte de um ano, e no horizonte de cinco anos. Com estes parâmetros, pode-se facilmente determinar o espaço requerido.

Após o cálculo do espaço requerido para armazenar os dados, faz-se o cálculo do espaço necessário para o armazenamento dos índices.

O resultado final obtém-se integrando estes resultados intermediários, e dará uma boa idéia da ordem de magnitude do DASD<sup>5</sup> necessário.

Finalmente, dispondo do resultado desta estimativa, é possível definir que enfoque de projeto e desenvolvimento deverá ser adotado.

Para o horizonte de um ano, se a perspectiva for de um total de 10.000 registros ou menos, praticamente qualquer técnica de projeto e implementação funcionará. Ainda para o mesmo horizonte, se o total for da ordem de 100.000 registros, o projeto deverá ser conduzido cautelosamente. Se a estimativa para o primeiro ano ultrapassar 1.000.000 de registros, níveis duais de granularidade se farão necessários. E caso a estimativa supere os 10.000.000 de registros no primeiro ano, níveis duais de granularidade serão obrigatórios, e tanto o projeto quanto a implementação deverão ser conduzidos com muita cautela (INMON, 1997).

Para o horizonte de cinco anos, os totais se alteram em aproximadamente uma ordem de grandeza. A teoria indica que (INMON, 1997):

---

<sup>5</sup> Direct access storage device

- Haverá maior disponibilidade de conhecimento sobre gerenciamento de grandes volumes de dados no *Datawarehouse*.
- Os custos de hardware terão diminuído consideravelmente.
- Ferramentas de software mais poderosas estarão disponíveis.
- Usuário final será mais sofisticado.

Feita esta análise, o próximo passo consiste em definir exatamente qual será o nível de granularidade a ser adotado. Aí vale muito a experiência do desenvolvedor.

De qualquer maneira, é recomendável utilizar um processo iterativo com o usuário final para se ajustar corretamente a granularidade que será utilizada.

### **Particionamento**

Outro tópico que também deve ser considerado no ambiente do *Datawarehouse* é o Particionamento (INMON, 1997), isto é, como os dados de detalhe deverão ser particionados.

O objetivo do particionamento dos dados de detalhe é repartir estes dados em unidades físicas menores e, desta forma, dar maior flexibilidade para o gerenciamento dos mesmos (ZIMMER, 2001), que é uma das características essenciais do *Datawarehouse*.

Ao serem particionados, os dados afins devem ser colocados todos juntos numa mesma partição. Vários são os critérios que podem ser utilizados:

- Por data
- Por negócio
- Por área geográfica
- Por unidade organizacional
- Por todos os critérios acima



Devido a periodicidade dos dados, um dos critérios praticamente obrigatório é o critério da data.

Uma das questões mais desafiadoras para o desenvolvedor do *Datawarehouse*, é saber se faz o particionamento no nível de sistema, ou no nível de aplicação (INMON, 1997). O particionamento no nível de sistema é feito diretamente no DBMS, enquanto no nível de aplicação, é controlado pelo desenvolvedor que atua diretamente no código fonte da aplicação; neste caso, nem o DBMS nem o sistema operacional tomam conhecimento de qualquer relação existente entre as partições.

Como regra, é mais conveniente particionar os dados no nível de aplicação, e talvez a razão mais forte para isto, seja a possibilidade de fazer uma definição de dados diferente para cada ano.

Quando o particionamento é feito no nível de sistema, o DBMS inevitavelmente assume uma definição de dados única. Considerando que o *Datawarehouse* mantém dados referentes a um período longo de tempo, até dez anos, e considerando também que a definição dos dados normalmente é alterada, não faz sentido permitir que o DBMS imponha a característica de definição única para os dados.

Outro fator importante para justificar que o particionamento seja feito no nível de aplicação, é a possibilidade de transferência da partição de um complexo de processamento para outro. Quando a carga de trabalho e o volume de dados ficam pesados para o ambiente do *Datawarehouse*, este fator se mostra como uma vantagem concreta.

O teste final do particionamento de dados ocorre quando é feita a pergunta: “É possível acrescentar um índice a uma partição sem que qualquer interrupção seja sentida pelas outras operações?” (INMON, 1997).

Se houver a possibilidade de acrescentar um índice livremente, significa que o particionamento está satisfatório. Caso contrário, a partição deverá ser dividida de maneira mais adequada.

### **3.3 Datamining**

#### **3.3.1 Conceito**

A maioria das organizações vem acumulado uma enorme quantidade de dados ao longo do tempo, mas na realidade elas precisam mesmo é de informações.

Informações que possam ser obtidas a partir destes dados, e que sejam utilizadas para atender cada vez melhor os seus clientes, para alocar os seus recursos de uma maneira mais eficiente, para minimizar perdas, para subsidiar a tomada de decisões, etc.

Mas, como extrair as informações necessárias? Uma tecnologia nova, emergente nos anos 90, tem o propósito de achar a solução. Esta tecnologia é chamada de *Datamining*, e usa sofisticadas técnicas de modelagem e de análises estatísticas, para descobrir padrões e relacionamentos escondidos em grandes bases de dados (BRACHMAN e ANAND, 1996), (FAYYAD et al. 1996).

*Datamining*, ou mineração de dados, é o processo de extrair informação válida, previamente desconhecida e de máxima abrangência a partir de grandes bases de dados, usando-as para efetuar decisões cruciais. *Datamining* vai muito além da simples consulta a um banco de dados, pois permite aos usuários explorar e inferir informação útil a partir dos dados, e descobrir relacionamentos escondidos no banco de dados. Pode ser considerada uma forma de descobrimento de conhecimento em bancos de dados (*KDD - Knowledge Discovery in Databases*) (FAYYAD et al. 1996), área de pesquisa

de bastante evidência no momento, envolvendo Inteligência Artificial e Banco de Dados.

*Datamining* descende fundamentalmente ([www.businessobjects.com](http://www.businessobjects.com)) de 3 linhagens. A mais antiga delas é a Estatística Clássica. Sem a estatística não seria possível termos o *Datamining*, visto que a mesma é a base da maioria das tecnologias a partir das quais o *Datamining* é construído.

A Estatística Clássica envolve conceitos como distribuição normal, variância, análise de regressão, desvio simples, análise de conjuntos, análises de discriminantes e intervalos de confiança, todos usados para estudar dados e os relacionamentos entre eles.

Estas são as pedras fundamentais onde as mais avançadas análises estatísticas se apóiam, e, sem dúvida, no coração das atuais ferramentas e técnicas de *Datamining*, a análise estatística clássica desempenha um papel fundamental.

A segunda linhagem do *Datamining* ([www.businessobjects.com](http://www.businessobjects.com)) é a Inteligência Artificial. Essa disciplina, que é construída a partir dos fundamentos da heurística, em oposto à estatística, tenta imitar a maneira como o homem pensa na resolução dos problemas estatísticos. Em função desta abordagem, ela requer um impressionante poder de processamento, que era impraticável até os anos 80, pois foi a partir daí que os computadores começaram a oferecer um bom poder de processamento, à preços cada vez mais acessíveis.

A Inteligência Artificial desenvolveu algumas aplicações para o alto escalão do governo/cientistas americanos, sendo que os altos preços não permitiram que ela ficasse ao alcance de todos. As notáveis exceções foram certamente alguns conceitos de Inteligência Artificial adotados por alguns produtos de ponta, como módulos de otimização de consultas para Sistemas de Gerenciamento de Banco de Dados..

A terceira e última linhagem do *Datamining* ([www.businessobjects.com](http://www.businessobjects.com)) é a chamada *Machine Learning*, que pode ser melhor descrita como o casamento entre a Estatística Clássica e a Inteligência Artificial. Enquanto a Inteligência Artificial não se transformava em sucesso comercial, suas técnicas foram sendo largamente utilizadas pela *Machine Learning*, que foi capaz de se valer das sempre crescentes taxas de preço/performance oferecidas pelos computadores nos anos 80 e 90, conseguindo mais e mais aplicações devido às suas combinações entre heurística e análise estatística. A *Machine Learning* (QUINLAN, 1993) tenta fazer com que os programas de computador “aprendam” com os dados que eles estudam, tal que esses programas tomem decisões diferentes baseadas nas características dos dados estudados, usando a estatística para os conceitos fundamentais, e adicionando algoritmos heurísticos avançados, pertencentes à Inteligência Artificial, para alcançar os seus objetivos.

*Datamining* é fundamentalmente ([www.businessobjects.com](http://www.businessobjects.com)) a adaptação das técnicas da *Machine Learning* para as aplicações em negócios. Desse modo, podemos descrevê-lo como a união dos históricos e dos recentes desenvolvimentos em Estatística, em Inteligência Artificial e *Machine Learning*. Essas técnicas são usadas juntas para estudar os dados e achar tendências e padrões nos mesmos.

*Datamining* tem experimentado uma crescente aceitação nas ciências e nos negócios, notadamente aqueles que precisam analisar grandes volumes de dados e achar tendências que não poderiam ser descobertas de outra forma.

*Datamining* descobre estes padrões e relacionamentos através da construção de modelos, que são representações abstratas da realidade. Um mapa, por exemplo, é um modelo; nele é possível ver o caminho desde o aeroporto até o centro da cidade, mas

não é apresentada uma possível situação de trânsito lento provocada por um acidente, ou um desvio provocado por uma obra.

Isto quer dizer que, embora nunca se deva confundir o modelo com a realidade, um bom modelo é sempre um guia muito útil para entender o negócio da organização e sugerir ações que o melhorem.

Existem dois tipos de modelos na tecnologia de *Datamining* (BERRY e LINOFF, 1997).

O primeiro, chamado de **Modelo Probabilístico**, usa dados e resultados conhecidos, para desenvolver modelos que possam prever resultados a partir de diferentes dados. Por exemplo, pode ser construído um modelo usando dados históricos de pessoas para as quais foram concedidos empréstimos, com a finalidade de ajudar a identificar pessoas para as quais não se devem conceder empréstimos.

O segundo, chamado de **Modelo Descritivo**, descobre os padrões existentes nos dados e utiliza-os para subsidiar a tomada de decisões. A diferença fundamental entre os dois tipos de modelos é que o modelo probabilístico faz previsões explícitas, tais como o lucro esperado, o número de inadimplentes, etc., enquanto o modelo descritivo é usado para ajudar a construir o modelo probabilístico, ou para fazer previsões implícitas que formam a base para uma ação ou decisão.

Evidentemente, toda empresa que conhece o seu negócio e os seus clientes, está sempre bem informada sobre os padrões mais significativos que foram descobertos ao longo do tempo. O que o *Datamining* pode fazer, não é apenas confirmar estas observações empíricas, mas também descobrir novos padrões, alguns até muito sutis e difíceis de serem observados empiricamente (BRACHMAN e ANAND, 1996). Estes novos conhecimentos podem trazer grandes retornos para a empresa, pois propiciam um

melhoramento contínuo. Assim, obtém-se uma pequena vantagem a cada mês, a cada projeto, a cada cliente, vantagem esta que contabilizada num período maior de tempo, faz o diferencial competitivo em relação às empresas que não utilizam bem o *Datamining*.

É prudente lembrar, no entanto, que *Datamining* é uma ferramenta e nada mais que uma ferramenta. Desta forma, ele não vai tomar a iniciativa de ficar vigiando o que acontece na sua fonte de dados, nem lhe mandar um e-mail quando descobrir um padrão interessante, ou seja, ele não elimina a necessidade de conhecer o seu negócio, de entender os seus dados, nem de estar familiarizado com os métodos analíticos utilizados.

*Datamining* ajuda o analista de negócios a descobrir padrões e relacionamentos entre os seus dados. Como são utilizados modelos, os padrões e relacionamentos descobertos pelo *Datamining* devem ser comprovados no mundo real.

*Datamining* requer o conhecimento das ferramentas utilizadas e dos algoritmos nos quais elas se baseiam, pois eles são diretamente relacionados com a precisão e velocidade obtidas pelo modelo.

Requer também um bom entendimento dos dados, pois a qualidade dos resultados obtidos pelos algoritmos é sensível aos *outliers* (dados que são discrepantes dos demais dados pertencentes à fonte de dados), aos atributos irrelevantes ou atributos que variam juntos, (como idade e data de nascimento, por exemplo), à maneira como os dados foram codificados, etc.

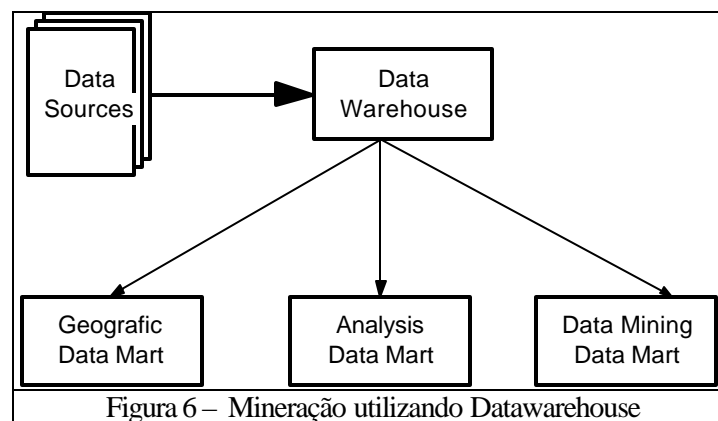
Seria enganoso dizer que *Datamining* responderá perguntas que não foram feitas. Embora não seja necessário perguntar (“Será que os franceses entre 18 e 25 anos terão interesse no produto que anunciei ?”), ainda será necessário solicitar ao *Datamining*

que descubra o padrão que se está procurando. Isto poderia ser feito de uma maneira bem vaga, como (“Ajude a descobrir quem responderá ao meu anúncio”). Provavelmente as duas respostas seriam diferentes.

*Datamining* não substitui o analista de negócios ou o gerente experiente, ao invés disto dá a eles uma poderosa ferramenta para melhorar o trabalho que executam.

### 3.3.2 Datamining e Datawarehouse

Os dados que vão ser “minerados”, freqüentemente são extraídos do *Datawarehouse* da organização, conforme mostra a Figura 6, onde *Data Mart* é um sub-conjunto do *Datawarehouse*, orientado para um assunto específico ou filtrado por área. (por departamento, como exemplo).



Existem benefícios reais ao se utilizar esta abordagem. Os problemas de refinamento dos dados para *Datawarehouse* e para *Datamining* são similares. Daí se os dados forem pesquisados diretamente no *Datawarehouse*, muitos dos problemas envolvidos com a sua consolidação já terão sido resolvidos.

Todavia, *Datawarehouse* não é um pré-requisito para *Datamining*, como mostra a Figura 7, onde a mineração dos dados é feita diretamente a partir de bases de dados operacionais.

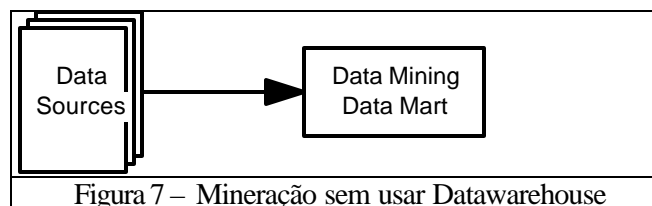


Figura 7 – Mineração sem usar Datawarehouse

Às vezes se utiliza esta abordagem, para o evitar a construção de um *Datawarehouse*, que pode exigir um tempo muito grande e envolver um custo muito alto.

A tendência universal nítida é a utilização das duas tecnologias de uma maneira harmônica, onde o *Datawarehouse* faz todo o trabalho relativo à preparação e disponibilização adequada dos dados, que posteriormente serão utilizados pelo *Datamining* para produção das informações de interesse.

### 3.3.3 Classificação dos Métodos

Diferentes esquemas podem ser utilizados para classificar os métodos dentro da tecnologia de *Datamining*, e eles levam em conta (CHEN , 1996):

- tipo de base de dados que será utilizado.

É considerado um *Datamining* relacional, o sistema que utiliza uma base de dados relacional para extrair os conhecimentos, da mesma forma que é considerado um *Datamining* orientado a objeto, aquele que extrai conhecimentos de uma base de dados orientada a objetos.



Qualquer tipo de base de dados pode ser utilizada, e o *Datamining* é classificado em função dela: transacional, espacial, temporal, multimídia, informações da Internet, entre outras.

- tipo de conhecimento a ser extraído.

Muitos são os tipos de conhecimentos possíveis de serem extraídos utilizando-se *Datamining*, incluindo entre eles as regras de associação, regras de classificação, regras discriminantes, *clustering*, evolução, análise de desvio, etc.

É possível ainda classificar o *Datamining* de acordo com o nível de abstração do conhecimento descoberto, que pode ser genérico, específico ou até múltiplo.

- tipo de técnica a ser utilizada.

Outra classificação pode ser obtida considerando o método utilizado para minerar o conhecimento; através de *query*, usando interatividade, etc., ou considerando a abordagem, que pode ser baseada em teorias estatísticas, em padrões, em generalizações, ou mesmo uma abordagem integrada.

### 3.3.4 Técnicas Utilizadas

Existem várias técnicas para extração de conhecimento e mineração de dados; elas podem ser divididas em três grupos: consulta/visualização, classificação e *clustering*, como veremos a seguir (BERRY e LINOFF, 1997).

#### Grupo Consulta/Visualização

*Datamining* é baseada em inteligência artificial e as consultas submetidas aos seus algoritmos não precisam ser tão objetivas como aquelas submetidas a modelos de previsão de tendências baseados em dados do passado. É muito útil para extrair informações que são previamente desconhecidas (BERSON, 1997).

*OLAP-On-line Analytical Processing e Decision Support Systems – DSS* por outro lado, consideradas técnicas de consulta e visualização, dependem da interação humana para formular bem a questão, e esta é a sua maior desvantagem. (BERSON, 1997).

Uma questão típica de *OLAP* ou *DSS* poderia ser: “As pessoas da região nordeste do Brasil compram mais sorvete que as da região sul do Brasil?”.

A resposta é obtida de uma análise estatística pura.

No caso do *datamining*, a questão poderia ser: “Descubra o perfil que identifica as características mais prováveis das pessoas que compram sorvete”.

A resposta poderia indicar que as vendas dependem da estação do ano e das condições climáticas.

Desta forma vê-se que *datamining* pode descobrir fatos sutis ou até mesmo sofisticados que extrapolam os limites do *OLAP / DSS*.

### **Grupo Classificação – Regras de Associação**

Classificação de dados é o processo de descoberta de propriedades comuns a um conjunto de objetos de uma base de dados, com o objetivo de enquadrá-los em determinadas classes, de acordo com um modelo de classificação.

Para a construção de um modelo de classificação parte-se de um **conjunto de treinamento**, que é uma base de dados de pequena dimensão cujas classes já são conhecidas, e cujos atributos são os mesmos existentes na base de dados real que vai ser utilizada.

O objetivo da classificação é analisar os dados do conjunto de treinamento e desenvolver uma descrição precisa (modelo) para cada classe, usando as características conhecidas dos dados. Estes modelos são então utilizados para classificar dados na base

de dados real ou para desenvolver uma melhor descrição, chamada regra de classificação para cada classe existente na base de dados.

A classificação de dados é utilizada em aplicações de diagnóstico médico, previsão de tendências, determinação de estratégias de *marketing*, etc., e vem sendo estudada em estatística, *machine learning*, redes neuronais e sistemas especialistas (WEISS e KULIKOWSKI, 1991) e é um importante tema em *datamining* (FAYYAD et al. 1996).

As **regras de associação** têm por objetivo descobrir importantes associações entre itens que compõem uma base de dados, de tal forma que a presença de um item numa determinada transação, implique na presença de outro item na mesma transação.

O modelo matemático está descrito a seguir:

Seja  $I = \{ i_1, i_2, \dots, i_m \}$  um conjunto de itens. Seja  $D$  um conjunto de transações, onde cada transação  $T$  é um conjunto de itens, tal que  $T \subseteq I$ . Note que a quantidade de itens da transação não é considerada, mas sim a sua existência. Cada transação é associada com um identificador chamado **TID**. Seja  $X$  um conjunto de itens. Uma transação  $T$  contém  $X$  se e somente se  $X \subseteq T$ .

Uma regra de associação é uma implicação da forma  $X \Rightarrow Y$ , onde  $X \subseteq I$ ,  $Y \subseteq I$  e  $X \cap Y = \emptyset$ .

A regra  $X \Rightarrow Y$  sobre o conjunto de transações  $D$  apresenta uma confiabilidade  $c$  se  $c\%$  das transações em  $D$  que contêm  $X$  também contêm  $Y$ .

A regra  $X \Rightarrow Y$  sobre o conjunto de transações  $D$  apresenta um suporte  $s$  se  $s\%$  das transações em  $D$  contêm  $X \cup Y$ .

Confiabilidade indica a firmeza da implicação, ao passo em que suporte indica a frequência de ocorrência dos padrões que estão sendo examinados na regra.

É recomendável trabalhar apenas com as regras que têm um suporte significativo.

Regras que possuem alta confiabilidade e grande suporte são chamadas de ***strong rules***

(AGRAWAL, IMIELINSKI e SWAMI, 1993) e (PIATETSKY-SHAPIRO, 1991).

Em (AGRAWAL, IMIELINSKI e SWAMI, 1993), (AGRAWAL e SRIKANT, 1994) e (PARK, CHEN e YU, 1995), o problema de descoberta de regras de associação é decomposto em duas etapas:

Descobrir os grandes conjuntos de itens, isto é, os conjuntos de itens que têm suporte nas transações acima de um valor pré-determinado.

Usar estes conjuntos de itens para gerar as regras de associação existentes na base de dados.

A performance obtida na geração das regras de associação é determinada pela primeira etapa, pois exige um processamento de dados muito grande, após o qual a geração das regras propriamente ditas pode ser feita de uma maneira razoavelmente mais simples. Desta forma os algoritmos existentes se propõem a resolver eficientemente o problema da primeira etapa.

Para ilustrar, são apresentados os algoritmos **Apriori** e **DHP** desenvolvidos em (AGRAWAL e SRIKANT, 1994) e (PARK, CHEN e YU, 1995),

### **Algoritmos Apriori e DHP**

Considere o exemplo de transação numa base de dados apresentado na Tabela abaixo:

| TID | Itens |
|-----|-------|
| 100 | A C D |
| 200 | B C E |

|     |         |
|-----|---------|
| 300 | A B C E |
| 400 | B E     |

Em cada iteração, **Apriori** constrói um conjunto itens candidatos, conta o número de ocorrências de cada item e seleciona o conjunto baseado num suporte mínimo predeterminado. Na primeira iteração, o algoritmo simplesmente percorre todas as transações para contar o número de ocorrências de cada item, obtendo assim o resultado apresentado na tabela a seguir.

Candidatos de 1 Item

| Item | Suporte |
|------|---------|
| {A}  | 2       |
| {B}  | 3       |
| {C}  | 3       |
| {D}  | 1       |
| {E}  | 3       |

Supondo que o suporte mínimo requerido seja 2 (40% dos itens), o item D é imediatamente eliminado.

Para descobrir o conjunto de candidatos composto por 2 itens, o algoritmo combina os candidatos de 1 item aprovados, e percorre então todas as transações deste novo conjunto para contar as ocorrências, conforme apresentado na tabela a seguir:

Candidatos de 2 Itens

| Item  | Suporte |
|-------|---------|
| {A B} | 1       |

|       |   |
|-------|---|
| {A C} | 2 |
| {A E} | 1 |
| {B C} | 2 |
| {B E} | 3 |
| {C E} | 2 |

Levando-se em conta o suporte, os itens {A B} e {A E} são também eliminados ( $s < 2$ ).

Para descobrir os candidatos compostos por 3 itens, o algoritmo adota procedimento análogo ao mostrado anteriormente, utilizando como base o conjunto de candidatos de 2 itens que foram aprovados, para obter finalmente o resultado apresentado abaixo:

Candidatos de 3 Itens

| Item    | Suporte |
|---------|---------|
| {B C E} | 2       |

Da mesma maneira que o **Apriori**, o algoritmo **DHP** também gera o conjunto de candidatos de  $k$  itens, partindo do conjunto de candidatos de  $k-1$  itens. A diferença é que o **DHP** utiliza uma tabela de *hashing*, que é construída no passo anterior, para testar a elegibilidade do candidato, e, devido à este artifício., só coloca no novo conjunto, os candidatos cujo suporte é maior ou igual ao suporte mínimo requerido.

Pode-se dizer que o algoritmo **DHP** é um melhoramento do algoritmo **Apriori**, pois apresenta um desempenho superior, diretamente relacionado com a diminuição das complexidades memória e tempo.

### **Grupo Classificação - Regras de Associação em Múltiplos Níveis.**

Existem situações onde as associações interessantes só são percebidas num nível de abstração muito alto. Por exemplo, utilizando-se uma base de dados transacional de um supermercado, pode-se obter um padrão de compras relacionando pão com leite, e é possível que não se obtenha nenhuma regularidade ao se tentar relacionar estes mesmos itens com maiores níveis de detalhe, como leite desnatado longa vida Tirol e pão integral Seven Boys, por exemplo.

Devido a este fato, é importante estudar as regras de associação num nível generalizado de abstração (SRIKANT e AGRAWAL, 1995) ou em multi-níveis (HAN et al. 1995).

As fontes de dados podem ser preparadas com múltiplos níveis de abstração, e desta forma facilitar o processo de geração de regras de associação. Por exemplo, numa hierarquia de classes os atributos podem estar expressos em dias, sintetizados em meses e ainda sintetizados em anos.

As associações num nível mais baixo só devem ser examinadas quando suas associações correspondentes em níveis mais altos forem significativas; observe-se que suportes mínimos diferentes devem ser adotados para níveis diferentes.

Em (SRIKANT e AGRAWAL, 1995) são estudados métodos para descobrir regras de associação num nível generalizado de abstração, utilizando uma extensão do algoritmo **Apriori**, apresentado anteriormente.

### **Significância das Regras de Associação Descobertas**

Nem todas as regras de associação descobertas são significativas o suficiente para serem utilizadas.

Por exemplo, considere-se o caso obtido pela mineração do resultado de uma pesquisa feita com 5.000 estudantes:

60% deles (3.000 estudantes) jogam basquete, 75% deles (3.750) comem cereal e 40% deles (2.000 estudantes) jogam basquete e comem cereal.

Suponha que foram fornecidos os seguintes parâmetros para serem utilizados na descoberta de regras de associação: suporte mínimo de 2.000 estudantes e confiabilidade mínima de 60%.

A seguinte regra de associação será produzida: **joga basquete → come cereal** já que ela satisfaz o suporte mínimo (2.000) e a correspondente confiabilidade:  $2.000 / 3.000 = 0.66$ ; apesar disto, esta regra é enganosa, pois a percentagem total de alunos que comem cereal é 75%, maior que 66%.

Como este fato pode conduzir a erros na tomada de decisões, pode-se dizer que uma regra de associação **A → B** só é interessante se sua confiabilidade exceder uma certa medida, ou seja: Se  $( (P(A \text{ e } B) / P(A)) - P(B) ) > d$ , onde  $d$  é uma constante escolhida convenientemente.

Esta expressão representa um teste de independência estatística, e deve ser levada em consideração para determinar se as regras descobertas devem ou não ser utilizadas.

### **Grupo Classificação - Generalização de dados.**

As bases de dados nos seus níveis primitivos contêm detalhes, como por exemplo, código do item, nome, descrição, fabricante, data de fabricação, preço unitário de compra, data de compra, unidade de medida, etc., e usualmente se deseja sintetizar conjuntos de dados relacionados, para fazer uma descrição generalizada; este processo é conhecido como generalização de dados.

Os métodos utilizados para fazer a generalização de uma maneira eficiente e flexível podem adotar duas abordagens: *Cubo de Dados* e *Indução Orientada a Atributos*.



### **Abordagem do Cubo de dados ( *Data Cube* )**

A abordagem do Cubo de Dados (*Data Cube*), discutida em (GUPTA et al. 1995, HARINARAYAN et al. 1996, WIDOM, 1995 e YAN et al. 1995), também é referenciada como *Base de Dados Multidimensional*, *Visões* e *OLAP → On-line Analytical Processing*.

A idéia básica desta abordagem é preparar previamente certas consultas que são freqüentemente solicitadas e que envolvem demorados processamentos, especialmente aqueles relacionados à agregação de funções, tais como quantidade, soma, média, mínimo, máximo, etc., e armazenar estes resultados (visões) numa base de dados multidimensional, chamada de *Data Cube*, para utilizá-los em aplicações que dão suporte à tomada de decisões, descoberta de novos conhecimentos, etc.

### **Abordagem da Indução Orientada a Atributos**

A abordagem da Indução Orientada a Atributos (HAN et al. 1993, HAN e FU, 1996), leva em conta a possibilidade dos dados contidos no *datawarehouse* não refletirem os dados mais atuais da base de dados, devido a periodicidade utilizada na migração, e então, para a generalização dos dados, faz consultas utilizando SQL, consultas estas que extraem diretamente da base de dados os resultados de interesse.

O aspecto principal desta abordagem é a natureza *on-line* da generalização dos dados, que é feita primeiro examinando a distribuição dos dados para cada atributo do conjunto relevante, para em seguida fazer os cálculos necessários de acordo com os níveis de abstrações desejados e finalmente obter as correspondentes generalizações.

### **Grupo Classificação - Árvores de Decisão**

O método de classificação baseado em árvores de decisão (QUINLAN, 1986 e QUINLAN, 1993) é um método de aprendizado supervisionado, que constrói árvores

de decisão a partir de um conjunto de exemplos conhecidos. É muito utilizado em *machine learning*.

O método inicia escolhendo um subconjunto do conjunto de treinamento (uma janela) e montando uma árvore com os elementos desta janela. Se esta árvore não produzir uma resposta correta para todos os elementos, exceções vão sendo inseridas na janela, e o processo continua até que se encontre uma árvore de decisão correta. O resultado final é uma árvore onde cada folha contém um nome de classe e cada nó especifica um atributo com os correspondentes desvios para cada valor possível deste atributo.

Um sistema de aprendizado típico baseado em árvore de decisão ID-3 (QUINLAN, 1986), adota a estratégia *top-down*, que pesquisa apenas parte do espaço da busca; esta estratégia garante encontrar rapidamente uma solução simples, não necessariamente a mais simples.

O sistema ID-3 para fazer a classificação dos objetos, usa o critério do ganho de informação, que é expresso na sua função de avaliação

$$i = \sum (p_i \ln(p_i))$$

onde  $p_i$  é a probabilidade de um objeto pertencer a classe  $i$ .

Existem diversas funções de avaliação como *Gini index*, *chisquare test* e outras que podem ser encontradas em (BREIMAN et al. 1984, KLOSGEN, 1996, PIATETSKY-SHAPIRO, 1991, WEISS e KULIKOWSKI, 1991), como também existem abordagens para transformar árvores de decisão em regras (QUINLAN, 1993), e regras em árvores de decisão (GAINS, 1996).

### **Grupo Clustering**

*Clustering* é um processo de agrupamento de objetos utilizando as suas características de similaridade. É feito de uma maneira não supervisionada, isto é, exige pouca ou

nenhuma intervenção do usuário, para encontrar as partições válidas existentes em grandes grupos de objetos.

Entre os diversos métodos de formação de *clusters* são muito bem aceitos os **estatísticos**, baseados em definições de distância para agrupar variáveis (hierárquicos), e para formar grupo de itens (não hierárquicos), e os de **redes neuronais artificiais** utilizados em aprendizagem não supervisionada.

Reveste-se de uma grande importância na solução de problemas onde *a priori* não se tem um bom conhecimento dos dados ou dos seus relacionamentos.

Existem vários algoritmos utilizados pelos métodos de formação de clusters, podendo ser destacados entre eles o K-Mean, descrito em (JOHNSON, 1998), que se enquadra nos métodos estatísticos e a rede de Kohonen, descrito em (KOHONEN, 1995), pertencente aos métodos de redes neuronais artificiais.

### 3.3.5 Métodos para Melhoria de Desempenho

O desempenho sempre foi o problema encontrado pela maioria das técnicas que atuam em grandes bases de dados. Elas funcionam razoavelmente bem em pequenas bases de dados, mas ficam sacrificadas em termos de desempenho, que se reflete ou no tempo de processamento, ou na qualidade da classificação obtida, quando trabalham com grandes bases de dados.

Por exemplo, o classificador de intervalos proposto por (AGRAWAL et al. 1992), cria índices na base de dados com a finalidade de melhorar apenas o desempenho da recuperação dos dados, e não a eficiência da classificação propriamente dita, tarefa esta que fica à cargo de algoritmo específico, o ID-3.

Uma integração direta da abordagem da indução orientada a atributos com o algoritmo ID-3, pode ajudar a descobrir regras em níveis mais altos de abstração; esta integração

aumenta a eficiência, mas pode reduzir a precisão da classificação obtida, dependendo do nível da generalização que foi feita.

Uma técnica de classificação multi-nível e de ajuste de níveis foi desenvolvida por (HAN et al. 1996), para melhorar a precisão do resultado da classificação em grandes base de dados, através da integração da abordagem da indução orientada a atributos com métodos de classificação.

(MEHTA et al. 1996) desenvolveu um classificador de regras para atuar em grandes bases de dados, chamado SLIQ – Supervised Learning in QUEST, que utiliza árvore de decisão e pode trabalhar tanto com atributos numéricos como com atributos categóricos. Utiliza uma pré-escolha na fase de construção da árvore, que está integrada com a estratégia de expansão da árvore em largura, para permitir a classificação de arquivos de dados residentes em dispositivos de armazenamento secundário.

SLIQ a cada passo usa também um algoritmo de poda, para compactar a árvore que está sendo construída. A combinação destas técnicas habilita o SLIQ a trabalhar bem com grandes bases de dados, independente do número de classes, atributos ou exemplos.

### **3.4 Aplicações de datawarehouse e datamining**

Com o intuito de exemplificar a aplicabilidade das tecnologias aqui descritas, serão apresentados alguns casos reais onde a utilização delas resultou em sucesso:

- A Wal-Mart, uma das maiores redes de lojas dos Estados Unidos, procurando em seu *datawarehouse* alguma relação entre o volume de vendas e os dias da semana, identificou, através do seu software de *datamining*, uma relação aparentemente estranha, mas muito forte, entre as vendas de fraldas descartáveis e as de cerveja, nas sextas feiras. Após esta descoberta, os produtos foram colocados lado a lado e as vendas aumentaram consideravelmente.

Este é um dos exemplos mais divulgados na literatura.

- A rede varejista Lojas Brasileiras descobriu que dos seus 51.000 produtos mantidos em estoque, muitos serviam apenas para atrapalhar o trabalho de gerenciamento. A rede tem setenta lojas distribuídas pelo Brasil, e com a utilização conjunta das técnicas de *datawarehouse* e *datamining*, descobriu que produtos encalhados como guarda chuvas, sombrinhas e malhas de lã, estavam em lojas na região nordeste, onde chuva e frio são raros. Descobriu também que foram enviados para Santa Catarina, onde a tensão é 220V, produtos com sistema único de 110V. Desta maneira a rede conseguiu retirar de suas prateleiras os produtos que apenas dificultavam o gerenciamento e influenciavam negativamente o seu desempenho
- O Bank of America usou *datawarehouse* e *datamining* para selecionar entre seus 36 milhões de clientes, aqueles com menor risco de dar calote num empréstimo. A partir das informações obtidas, enviou cartas oferecendo linhas de crédito para os correntistas cujos filhos tivessem entre 18 e 21 anos e, portanto, precisassem de dinheiro para ajudar os filhos a comprar o próprio carro, uma casa ou arcar com os gastos da faculdade. Resultado: em três anos, o banco lucrou 30 milhões de dólares.
- O Banco Itaú, pioneiro no uso de *datawarehouse* no Brasil, costumava enviar mais de um milhão de malas diretas, para todos os correntistas. No máximo 2% deles respondiam às promoções. Hoje, o banco tem armazenada toda a movimentação financeira de seus 3 milhões de clientes nos últimos 18 meses. A análise desses dados permite que cartas sejam enviadas apenas a quem tem maior chance de responder. A taxa de retorno subiu para 30%. A conta do correio foi reduzida a um quinto.( [dw@datawarehouse.inf.br](mailto:dw@datawarehouse.inf.br))

- A Sprint, um dos líderes no mercado americano de telefonia de longa distância, desenvolveu, com base no seu armazém de dados, um método capaz de prever com 61% de segurança se um consumidor trocava de companhia telefônica dentro de um período de dois meses. Com um marketing agressivo, conseguiu evitar a deserção de 120.000 clientes e uma perda de 35 milhões de dólares em faturamento.  
([dw@datawarehouse.inf.br](mailto:dw@datawarehouse.inf.br))

- A Union Pacific é uma empresa que gerencia 2.000 trens por dia, correndo sobre 31.000 milhas, passando por 25 estados, e que resolveu consolidar suas bases de dados através de um *Datawarehouse*, visando ganhar competitividade no mercado. Todas as áreas da empresa deveriam poder acessar com grande rapidez e flexibilidade os mesmos dados, que deveriam ser consistentes e confiáveis. Por exemplo, o marketing deveria analisar taxas e preços para descobrir o preço mais competitivo; a área de desenvolvimento deveria usar alguns dos mesmos dados para desenvolver rotas, enquanto a área financeira usaria também os mesmos dados para analisar os lucros.

A empresa, para auxiliar a dirigir os seus negócios, já tinha desenvolvido e vinha utilizando diversas aplicações. O problema era que os dados estavam em centenas de servidores e em *main frames* espalhados pela organização, trabalhando em ambientes diferentes, executando processos diferentes e armazenando dados em formatos diferentes; devido a estes fatos, era grande a dificuldade que tinham os usuários para encontrar respostas as questões que envolvessem todos estes dados.

Se alguém fizesse uma consulta sobre a situação da manutenção, por exemplo, receberia um relatório contendo uma pilha enorme de papel. Analisar aquelas

informações consumia muito tempo, e, dependendo da fonte utilizada, poderia conduzir a resultados diferentes, comprometendo a credibilidade da resposta.

A empresa resolveu então investir na tecnologia de *Datawarehouse* e hoje é reconhecida por ter feito um ótimo trabalho, tendo um sofisticado e eficiente *Datawarehouse*. Agora, de qualquer lugar e a qualquer tempo, pode-se acessar informações para suporte à decisão.

### 3.5 Sistemas existentes

Devido à imensidão de fatos a serem investigados no campo de mineração de dados, vários protótipos e sistemas foram desenvolvidos e vem sendo utilizados para extrair informações a partir de grandes bases de dados.

A seguir segue uma breve descrição de alguns deles:

- AC2 – um conjunto de bibliotecas escritas em C/C++ que possibilitam incluir as funcionalidades do *datamining* diretamente em qualquer software que está sendo desenvolvido por um profissional da área. Está disponível como uma *shell* para Windows 3.1 / 95 / 98 / NT e Unix..
- ALICE d'ISoft – uma ferramenta poderosa e de fácil utilização para fazer mineração de dados. Usa árvore de decisão para explorar os dados, gera SQL e relatórios textuais e faz análise do tipo *What-If*.
- DATA SURVEYOR – ferramenta de *datamining* para usuários *experts*. Consiste de um conjunto de algoritmos que dá suporte a todos os passos do processo de *knowledge discovery*. Possibilita ao usuário a descoberta interativa do conhecimento.
- DATA ENGINE é um software para analisar dados usando *fuzzy technologies*, *neural networks*, e estatística convencional. Vem sendo utilizado com sucesso em

previsão, vendas, controle de qualidade, análise de processos e diagnósticos. Inclui componente gráfico para visualização de dados.

- IBM VISUALIZATION DATA EXPLORER é um software interativo que permite aos cientistas, engenheiros e analistas visualizarem os seus dados em 3D. O programa usa um paradigma de programação visual para facilitar a construção de programas de visualização
- INLEN – este projeto trata do desenvolvimento de um sistema baseado em raciocínio multi-tipo, para extrair conhecimento a partir de grandes bases de dados. O sistema ajuda o usuário a descobrir padrões, tendências, relacionamentos expressivos, regularidades conceituais ou numéricas, com também anomalias em grandes bases de dados.
- IRIS – é um protótipo que suporta análise visual de dados referenciados espacialmente. Produz automaticamente mapas temáticos de alta qualidade para dados estatísticos fornecidos pelo usuário. O sistema aplica conhecimentos genéricos sobre como melhor combinar e apresentar dados estatísticos em mapas temáticos, com o auxílio regras heurísticas em conjunto com metadados específicos da aplicação, usando para isto uma linguagem de representação de conhecimento orientada a objetos.
- QUEST é um sistema de *datamining* desenvolvido na *IBM Almaden Research Center* por AGRAWAL et al. (1996), que descobre vários tipos de conhecimento em grandes bases de dados, incluindo regras de associação, regras de classificação, análise de padrões, etc.



- KEFIR é um sistema de *knowledge discovery* desenvolvido na *GTE Labs* por PIATETSKY-SHAPIRO (1991 e MATHEUS et al. (1996), para análise preventiva utilizando dos dados da saúde.
- SKICAT é um sistema de *knowledge discovery* desenvolvido na *Jet Propulsion Laboratory*, que detecta e classifica automaticamente objetos no céu, utilizando dados de imagem obtidas pelo laboratório.
- DBMiner é um sistema de *datamining* relacional, desenvolvido na *Simon Fraser University* por HAN et al. (1993, 1995) e HAN e FU (1996), que descobre múltiplos tipos de regras em múltiplos níveis de abstração.
- IMACS é um sistema de *datamining*, desenvolvido na *AT&T Laboratory* por BRACHMAN e ANAND (1996), que usa sofisticadas técnicas de representação do conhecimento.
- IDEA é um sistema desenvolvido na *AT&T Laboratory* por SELFRIDGE et al. (1996), que faz exploração e análise interativa de dados.
- SYNTHETIC CLASSIFICATION DATA SETS PROGRAM – Uma importante maneira para testar os algoritmos *learning-from-example* é avaliar suas performances contra conjuntos de dados sintéticos bem conhecidos. O *Synthetic Classification Data Sets* (SCDS) foi criado para gerar dados sintéticos que são particularmente úteis para testar os algoritmos de *Knowledge Discovery from Database* (KDD).
- CLEMENTINE é um conjunto de ferramentas para *datamining*, que ganhou duas vezes do *UK Government's (Department of Trade & Industry)* o prêmio SMART por inovação. Suas aplicações incluem segmentação e perfis de clientes no mercado,

detecção de fraudes, previsão de vendas para companhia de utilidades e previsão de lucros para empresas varejistas.

### 3.6 Conclusão

Neste capítulo foi feita uma revisão bibliográfica, sintetizando o estado da arte das tecnologias existentes e que podem ser utilizadas na solução dos problemas de gestão administrativo-tributária. Foram abordadas:

- Organização de grandes massas de dados (*Datawarehouse*)
- Extração de informações a partir de grandes massas de dados (*Datamining*)
- Fracionamento setorial do problema (*Datamarts*)
- Análise de informações sob diferentes óticas (*OLAP*)

É importante ter em mente que tecnologias e ferramentas não fazem um *Datawarehouse*. Torna-se necessário fazer um bom planejamento. Deve-se começar por um esboço aceitável tanto para negócios quanto para tecnologias participantes, documentar a estratégia de suporte à decisão e desenvolver um modelo para construir o *Datawarehouse*.

É importante ter um sócio que entenda e saiba de seu negócio. Ele fará o que for necessário para que seu *Datawarehouse* tenha sucesso. Para isso, divida as metas e escolha uma aplicação inicial que seja pequena e gerenciável (ZIMMER, 2001).

Faça de seus usuários uma parte do time; eles ajudarão a explorar a tecnologia, entendendo a capacidade e limitação da mesma; quanto mais você se comunicar com seus usuários, mais eles se sentirão parte do time. Desta forma eles perceberão os benefícios do *Datawarehouse* e esperarão mais acesso e mais vantagens. Além disso, mantendo sempre contato com eles, ficará fácil gerenciar suas expectativas.

O *Datawarehouse* pode ser visto como o alicerce do *Datamining* (ARMSTRONG, 2000), e a sua construção é uma jornada sem fim.

Assim como as informações da empresa mudam, muda também o *Datawarehouse*, de forma a adicionar novas capacidades e novas rotas à informação.

## 4 O MODELO PROPOSTO

### 4.1 Introdução

Este capítulo trata especificamente do modelo que foi desenvolvido no trabalho, e tenta explicar tão realisticamente quando possível, o processo que ocorreu desde o primeiro contato com o problema, até o seu completo equacionamento e representação através de um modelo.

Três foram as abordagens utilizadas na concepção do modelo: A primeira delas, “*Learning by Studying*”, foco principal da área de P&D nas engenharias, já está incorporada à rotina diária do autor, que, impulsionado pela própria atividade profissional, depara-se também sistematicamente com as abordagens “*Learning by Doing*” e “*Learning by Using*”, todas elas empregadas aqui.

A semente deste trabalho foi plantada no último trimestre de 1997, quando a UFSC – Universidade Federal de Santa Catarina, iniciou uma parceria com o CIASC (Centro de Informática e Automação do Estado de Santa Catarina), para estudar e analisar o problema de evasão fiscal no Estado de Santa Catarina. Como produto desta parceria, as partes obtiveram um melhor conhecimento dos reais problemas que afligem a administração tributária estadual, pesquisaram e testaram alternativas tecnológicas de solução, e desenvolveram um protótipo de grande repercussão nacional, que forneceu valiosos subsídios para este trabalho.

## 4.2 Etapas do Modelo

A seguir estão apresentadas as etapas do modelo proposto por este trabalho, que surgiu da conjugação de estudos sistemáticos, com incansáveis pesquisas, muito trabalho e porquê não, um pouco de inspiração.

### 4.2.1 Planejamento estratégico da necessidade de informações

A primeira etapa, indubitavelmente, deverá ser a obtenção iterativa do conhecimento dos requisitos funcionais do órgão onde se irá aplicar o modelo proposto, no caso específico deste trabalho, as SEFs.

Como resultado de cuidadoso levantamento efetuado nos níveis estratégico, tático e operacional, pode-se obter uma visão inicial das necessidades globais da Secretaria da Fazenda em foco, visão esta que irá sendo melhorada gradativamente em função de *feed-back* recebido dos profissionais envolvidos.

Normalmente pode ser encontrado nas SEFs um documento formal intitulado “Acompanhamento das Ações Fazendárias”, que é muito valioso para a consolidação das informações obtidas neste levantamento.

Trata-se de um documento oficial, emitido mensalmente, assinado pelo Secretário de Estado da Fazenda e elaborado com a participação das Diretorias internas como também do Conselho Estadual de Contribuintes. Este documento contém os seguintes itens:

➤ Acompanhamento da Arrecadação Tributária – onde são apresentados os resultados mensais provenientes da arrecadação dos três tributos de competência estadual, relacionados abaixo e constante do anexo I – Sistema tributário nacional:

- ICMS
- IPVA
- Causa Mortis

- Acompanhamento das Transferências Federais – onde são apresentados os resultados mensais das transferências de arrecadações provenientes do:
  - Fundo de Ressarcimento IPI - Exportações
  - Fundo de Participação dos Estados – FPE
  - Imposto de Renda Retido na Fonte dos Funcionários Públicos Estaduais
- Acompanhamento da Receita Líquida Disponível – demonstrativo da apuração mensal da receita líquida disponível e sua relação com o PIB<sup>6</sup> estadual.
- Acompanhamento das Despesas com Pessoal – resumo das despesas mensais com pessoal e estatísticas pertinentes.
- Acompanhamento do Contencioso Administrativo Tributário – informações sobre a quantidade de novos processos, processos julgados em primeira e em segunda instância.
- Acompanhamento das Ações de Fiscalização – informações sobre a quantidade e valor das notificações expedidas e estatísticas pertinentes.
- Acompanhamento da Cobrança do Imposto em Atraso – relação contendo a situação da cobrança do imposto em atraso, referente as notificações integrais e parciais, parcelamentos, confissão espontânea, dívida ativa integral e parcelada.
- Acompanhamento da Capacitação de Recursos Humanos – relação dos treinamentos realizados.
- Obras em Andamento – situação das obras em andamento.
- Processos em Andamento para Licitação - relação dos processos a serem licitados.
- Processos em Licitação - relação dos processos em licitação.

---

<sup>6</sup> PIB – Produto interno bruto.

Munido destes materiais, passa-se a construir e consolidar o planejamento estratégico da necessidade de informações da organização. Apesar do interesse aqui estar voltado para uma parte específica das necessidades das SEFs, neste caso a administração tributária, o planejamento precisa ser global. Uma boa análise buscando o entendimento do negócio da organização, já faz parte da solução que se busca. Acreditando e usando esta estratégia, apresenta-se numa sequência logicamente encadeada, as entidades identificadas como integrantes deste planejamento global:

**Região Fiscal:** Representa as unidades setoriais de fiscalização das SEFs.

**Município:** Representa as cidades do Estado.

**Funcionário:** Identifica os servidores públicos do Estado, podendo assumir as funções de Fiscal, Delegado, Auditor, Diretor, Representante, etc.

**Representação:** Identifica o órgão representativo da Fazenda em cada município que não seja sede da região fiscal.

**Banco:** Instituição bancária cujas agências autorizadas arrecadam tributos para o Estado.

**Agência Bancária:** Estabelecimento bancário credenciado a arrecadar qualquer Receita Estadual.

**Tabela Receita:** Representa os diferentes tipos de tributo arrecadado pela Fazenda Estadual.

**Pessoa:** Entidade super-tipo das entidades Pessoa Física e Pessoa Jurídica, que representa empresas e pessoas que se relacionam com a Fazenda Estadual.

**Pessoa Física:** Entidade sub-tipo de Pessoa, representando as pessoas físicas que recolhem Receitas Estaduais.

**Pessoa Jurídica:** Entidade sub-tipo de Pessoa, representando as empresas que recolhem Receitas Estaduais.

**Veículo:** Entidade super-tipo de veículo terrestre, aéreo e aquático, cuja propriedade incide o imposto sobre a propriedade de veículos automotores (IPVA).

**Veículo Terrestre:** Entidade sub-tipo de veículo, representando veículos automotores, cujo tráfego ocorre em vias públicas.

**Veículo Aquático:** Entidade sub-tipo de veículo, representando as embarcações de esporte e recreio, transporte e carga de passageiro, pesca e serviços em geral.

**Veículo Aéreo:** Entidade sub-tipo de veículo, representando as aeronaves de propriedade privada sujeitas a tributação de IPVA.

**Classe:** Agrupamento de veículos automotores cujas características fazem com que a alíquota de incidência do IPVA seja a mesma.

**Marca:** Identifica o fabricante do veículo, dentro de cada classe.

**Grupo:** Linha de veículos cujas características fazem com que todos os veículos pertencentes a ela, possuam o mesmo valor venal (base de cálculo).

**Base de Cálculo IPVA:** Registra o valor venal do veículo, que serve de base para o cálculo do IPVA correspondente ao mês que o tributo deve ser recolhido.

**Contribuinte:** Corresponde a toda pessoa física ou jurídica que contribui para a arrecadação estadual.

**ICMS Arrecadado:** Representa os impostos sobre circulação de mercadorias, devidos pelos Contribuintes, geralmente com apuração normal.

**IPVA Arrecadado:** Corresponde ao imposto sobre a propriedade de veículos automotores, com incidência anual, arrecadado pelos respectivos proprietários.



**Dívida Ativa:** Representa os créditos do Estado resultante de infrações cometidas por Contribuintes, executados judicialmente.

**Dívida Ativa Arrecadada:** Corresponde as amortizações parciais e/ou totais de débitos lançados em dívida ativa, porém não parcelados.

**Notificação:** Representa os valores devidos pelos contribuintes, decorrentes de infrações cometidas.

**Notificação Arrecadada:** Corresponde as amortizações parciais e/ou totais de débitos notificados, porém não parcelados, devidos pelos contribuinte ao Estado.

**Tributo Arrecadado por Processo:** Corresponde ao recebimento de parcelas oriundas de parcelamento de dívida ativa, notificação, confissão espontânea, etc.

**Taxa:** Representa cada recolhimento de taxas efetuado pela população, em contrapartida de serviços prestados pelo Estado.

**Processo:** Conjunto de documentação necessária à tramitação oficial de assuntos de interesse da população do Estado.

**Giam:** Documento contendo informações sobre o Icms a ser recolhido pelos contribuintes, cujo objetivo é fazer com que o Estado tenha uma previsão da arrecadação do mês.

**Componente Apuração Declarado:** Valores correspondentes a cada item da apuração do Icms.

**Tabela Componente:** Representa os componentes do documento Guia de Informações de Apuração Mensal.

**ICMS Devido:** Representa o valor mensal devido pelos contribuintes, referente a cada código do Icms.

**Dar/Gate:** Documento através do qual são recolhidos na rede bancária, os tributos e outras receitas estaduais.

**Órgão:** Entidades governamentais no âmbito federal, estadual e municipal que se relacionam com a administração estadual

**Çontador:** Representa os dados do contador responsável pela escrita fiscal e/ou contábil do contribuinte.

**Receita Alienação:** Representa os recursos oriundos da alienação de bens móveis e imóveis, promovida pelo poder público.

**ITBI/ITCMD Arrecadado:** Corresponde ao imposto sobre a transmissão de bens imóveis e ao imposto causa-mortis e doação.

**Valor Arrecadado:** Corresponde aos diversos valores arrecadados pela rede bancária (principal, multas, juros e acréscimos) através do Dar.

**Resumo Diário da Receita:** Corresponde ao resumo fornecido pela agência bancária, referente as receitas do Estado recolhidas no dia, por aquela agência.

**Tabela Grupo Receita:** Agrupamento de receitas com características semelhantes.

**Movimento por Grupo:** Representa os valores e quantidades arrecadados por grupo de receita.

**Totalizador Parcial Arrecadado:** Documento que totaliza os valores das Dar para efeito de conferência

**Processo Alienação de Bens:** Conjunto de informações sobre a alienação de bens do Estado.

**Processo Denúncia Espontânea:** Conjunto de informações sobre a confissão feita espontaneamente pelo contribuinte, sobre débitos do mesmo com o Estado.

**AIR Arrecadado:** Corresponde ao adicional do imposto de renda devido a União sobre lucros e ganhos de bens de capital.

**Empreendimento:** Obra ou serviço onde se aplicam recursos oriundos de receitas vinculadas.

**Transferência Vinculada:** Corresponde as receitas provenientes de acordos e convênios, cujas características obrigam que a aplicação seja efetuada em empreendimentos específicos.

**Convênios:** Acordos firmados entre o Estado e demais órgãos municipais, estaduais, federais e/ou internacionais.

**Parcela Convênio:** Receita correspondente a cada parcela de convênio, liberada pelo Estado.

**Financiamento:** Contratos firmados pelo governo do Estado com o intuito de obter recursos, à título de empréstimos, para saldar compromissos financeiros.

**Parcela Financiamento:** Parcela correspondente a cada parcela do financiamento liberada para o Estado.

**Receita Extra Orçamentária:** Corresponde a todos os valores pertencentes a terceiros, cuja guarda temporária é confiada ao Estado.

**LFTC:** Corresponde as letras financeiras do tesouro estadual, emitidas para a obtenção de recursos, colocadas a venda no mercado financeiro.

**Declaração AIR:** Conjunto anual de informações sobre o adicional do imposto de renda devido à União, a ser recolhido pelo contribuinte.

**Período Referência:** Representa os períodos de referência da arrecadação estadual, representados por mês e ano.

**Dia Útil:** Corresponde aos dias úteis dos meses onde pode haver arrecadação.

**DAF:** Divisão administrativa fiscal, onde se encontram agrupados os contribuintes para efeitos de fiscalização.

**Atuação Fiscal:** Registra a forma e o tempo de atuação do fiscal junto a divisão administrativa fiscal.

**Fiscal:** Funcionário capacitado e autorizado a efetuar oficialmente atividades de fiscalização junto ao contribuinte,.

**Posto Fiscal:** Representação da SEF com o intuito de efetuar, em caráter permanente, atividades de fiscalização de mercadorias em transito.

**Vinculação Contribuinte DAF:** Registra o período em que o contribuinte esteve ou está vinculado a uma DAF.

**Regionalização:** Representa a forma de agrupar os municípios do Estado, de acordo com a visão de cada secretaria ou órgão do governo.

**Vinculação Município Região Fiscal:** Informações referentes ao período em que o município esteve ou encontra-se vinculado a uma região fiscal.

**Equipe Posto:** Corresponde aos grupos formados nos postos fiscais para trabalho ininterrupto por um determinado período de tempo

**Tabela Item Infração:** Tabela onde são classificadas todas as infrações cometidas pelo contribuinte.

**Tabela Sub Item Infração:** Representa a sub classificação da infração cometida pelo contribuinte.

**Legislação:** Representa o diploma legal onde está baseada a aplicação de penalidades fiscais, na fiscalização de tributos estaduais..

**Capitulação Específica Principal:** Fundamentação legal da aplicação de notificação, com relação ao tributo principal notificado.

**Capitulação Específica Multa:** Fundamentação legal da aplicação de notificação, com relação a multa aplicada.

**Capitulação Específica Juros:** Fundamentação legal da aplicação de notificação, com relação aos juros aplicados.

**Capitulação Específica Correção Monetária:** Fundamentação legal da aplicação de notificação, com relação à correção monetária aplicada.

**Vigência Capitulação Principal:** Representa cada período de vigência da legislação que rege a aplicação de notificação concernente ao tributo principal.

**Vigência Capitulação Multa:** Representa cada período de vigência da legislação que rege a aplicação de notificação concernente à multa

**Vigência Capitulação Juros:** Representa cada período de vigência da legislação que rege a aplicação de notificação concernente aos juros.

**Vigência Capitulação Correção Monetária:** Representa cada período de vigência da legislação que rege a aplicação de notificação concernente à correção monetária.

**Capitulação Padrão Principal:** Fundamentação legal referente ao tributo principal notificado, cujas características enquadram-se numa capitulação padrão.

**Capitulação Padrão Multa:** Fundamentação legal referente à multa notificada, cujas características enquadram-se numa capitulação padrão.

**Capitulação Padrão Juros:** Fundamentação legal referente aos juros notificados, cujas características enquadram-se numa capitulação padrão.

**Capitulação Padrão Correção Monetária:** Fundamentação legal referente à correção monetária notificada, cujas características enquadram-se numa capitulação padrão.

**Remessa Blocos:** Corresponde às informações referentes a cada remessa de blocos de notificações enviadas as regiões fiscais, para distribuição aos fiscais.

**Bloco Notificações:** Conjunto de formulários pré-impressos, numerados sequencialmente, destinados a notificações.

**Folha Cancelada:** Corresponde as folhas pré-impressas destinadas a notificações, e que por qualquer motivo foram inutilizadas.

**Contribuinte Eventual:** Pessoa física ou jurídica não pertencente ao cadastro de contribuintes da SEF, mas que por algum motivo foi notificado ao transportar mercadorias.

**Amortização:** Corresponde as amortizações parciais e/ou totais de débitos de notificações não parcelados.

**Parcelamento Notificação:** Representa os valores notificados que, em acordo entre o contribuinte e o fisco, serão pagos de forma parcelada.

**Parcela:** Representa cada parcela de pagamento das notificações que sofreram processo de parcelamento.

**Plano Parcelamento:** Representa o plano de parcelamento que o contribuinte optou, sendo que cada grupo de parcelas representa um percentual sobre o montante da dívida.

**Processo Contencioso:** Conjunto de informações sobre a reclamação do contribuinte sobre notificações emitidas contra ele.

**Contencioso Notificação:** Identifica as contestações do contribuinte, assim como o resultado das decisões do órgão julgados.

**Fase Processo Contencioso:** Identifica a fase em que se encontra o processo contencioso no órgão julgador.

**Moeda:** Identifica as várias moedas correntes e estáveis usadas no Brasil.

**Cotação:** Registra o valor de uma moeda em relação a outra.

**Tabela Grupo Atividade Econômica:** Agrupamento por afinidade, de atividades econômicas exercidas pelos contribuintes.

**Tabela Atividade Econômica:** Representa as atividades econômicas exercidas pelos contribuintes.

**Regime Especial:** Situação em que se encontra o contribuinte em relação ao ICMS, com referência a autorização dada pelo fisco para recolhimento em prazos especiais.

**Situação Regime Especial:** Representa a tabela de situações estabelecidas em lei, onde o contribuinte pode requerer regime especial de recolhimento de tributos.

**Ato Decisório:** Representa o veredicto do processo contencioso, proferido pelo conselho estadual de contribuintes.

**Ementário:** Tabelas de ementas padrão (resumos e decisões) em que o ato decisório pode ser classificado.

**Participação Fiscal:** Corresponde a participação do fiscal nas notificações emitidas.

**Produção Fiscal:** Corresponde a produção mensal do fiscal, medida em parcelas produzidas através da emissão de notificações.

**Parcela Recebida:** Corresponde as parcela recebidas mensalmente pelos fiscais, correspondentes as suas produções na emissão de notificações.

**Componente Apuração Real:** Valores correspondentes a cada item componente do ICMS levantados pelo fisco.

**ICMS Real Devido:** Representa o valor mensal devido pelo contribuinte, referente a cada código do ICMS levantado pelo fisco.

**Processo Parcelamento Notificação:** Conjunto de informações sobre o pedido de parcelamento da notificação, solicitado pelo contribuinte.

**Grupo Cargo:** Representa os agrupamento dos cargos do Estado, levando em consideração a afinidade entre eles e definidos em regulamento próprio.

**Cargo:** Representa os cargos existentes na SEF.

**Nível:** Corresponde aos vários níveis salariais existentes em cada cargo.

**Receita Tributária:** Corresponde ao resumo mensal da receita tributária, arrecadada por região fiscal.

**Componente Denúncia:** Corresponde aos valores que compõem o processo de denúncia , efetuado pelo contribuinte.

**Credenciamento Gráfico:** Representa as autorizações anualmente concedidas aos estabelecimento gráficos para a impressão de documentos fiscais.

**AIDF:** Corresponde as autorizações concedidas as empresas para imprimirem seus documento fiscais em estabelecimentos gráficos credenciados.

**Item AIDF:** Corresponde as autorizações contidas na AIDF, referentes a cada série/subsérie solicitadas pelo contribuinte.

**Estabelecimento:** Pessoa jurídica representando as empresas que contribuem para a receita estadual, principalmente o ICMS.

**Quadro Societário:** Corresponde à composição societária de um estabelecimento no início de suas atividades, bem como em qualquer momento que a sua composição venha ser alterada.

**Participação Societária:** Corresponde a cada participante do quadro societário de um estabelecimento.

**Administração Estabelecimento:** Representa as pessoas que em função dos cargos exercidos no estabelecimento, estão autorizadas a assinar ou assumir responsabilidades perante o fisco pela empresa.



**Grupo Econômico:** Representa grupos com participação majoritária na formação do capital de várias empresas, normalmente empresas com atividades correlatas.

**DIEF:** Dados sobre a declaração anual de informações econômico-fiscais, apresentada a SEF com base nos documentos e livros fiscais

**DIEF Empresa Normal:** Corresponde a Dief para as empresas não classificadas como microempresas.

**DIEF Microempresas:** Corresponde a Dief para as empresas classificadas como microempresas.

**Unidade Federação:** Representa as unidades federativas do Brasil (estados brasileiros).

**Tabela Componente Valor Adicional:** Corresponde aos itens que fazem parte da composição do valor acrescido pelas empresas (valor adicional).

**Tabela Componente Receita Bruta:** Corresponde aos itens que fazem parte da composição da receita bruta das empresas.

**Movimento Mensal Dief-ME:** Corresponde as entradas e saídas de mercadorias efetuadas mensalmente pelas microempresas.

**Compra Produtores Agropecuários/Pescadores:** Registra as compras de qualquer produto fornecido por produtores agropecuários e pescadores, distribuídas por município.

**Prestação Serviços/Fornecimento Energia Elétrica:** Registra as receitas oriundas da prestação de serviços sujeitos ao ICMS, e oriundas do fornecimento de energia elétrica, distribuídas por município.

**Operações Unidade Federação:** Informações sobre o movimento de entrada e saída de mercadorias entre o estado em consideração e os demais estados, efetuado pelas empresas normais.

**Movimento Entrada Saída:** Resumo das entradas e saídas de mercadorias, transcrito do livro de apuração do ICMS das empresas normais.

**Exclusão Valor Adicionado:** Valores correspondente as entradas e saídas que não são consideradas na apuração do valor adicionado a mercadoria (valor agregado).

**Componente ICMS Declarado DIEF:** Valores anuais correspondente aos débitos e créditos de ICMS e apuração do imposto devido no exercício.

**Apuração Receita Bruta:** Corresponde aos valores em moeda estável (ufir) do componente de apuração da receita bruta anual das microempresas.

**Comarca:** Região da estrutura judiciária do Estado.

**Processo Contencioso Dívida Ativa:** Conjunto de informações sobre a reclamação do contribuinte sobre a dívida associada a ele.

**Processo Adjudicação:** Conjunto de informações sobre a tomada judicial de bens, como pagamento de dívida ativa.

**Bem Adjudicado:** Representa os itens recebidos pelo Estado em pagamento de dívida ativa.

**Amortização Dívida Ativa:** Registra todo e qualquer pagamento, total ou parcial, da dívida ativa.

**Advogado:** Profissional de advocacia, que representa o Estado mediante credenciamento junto a vara da Fazenda.

**Participação Dívida Ativa:** Corresponde à participação do advogado naquela dívida, para efeito de distribuição dos honorários advocatícios.

**Participação Dívida Parcelada:** Corresponde à participação do advogado no parcelamento da dívida, para efeito da distribuição dos honorários advocatícios.

**Garantia Dívida:** Representa os itens oferecidos pelo devedor como garantia de pagamento da dívida.

**Atividade Econômica Exercida:** Corresponde as atividades econômicas exercidas pelos contribuintes da fazenda estadual.

**Documento Fiscal:** Conjunto de informações sobre o documento fiscal (Nota Fiscal, Selo, Romaneio, Ticket de Caixa, etc) que registra a entrada ou saída de mercadorias em estabelecimentos.

**Item Documento Fiscal:** Representa cada item de mercadoria ou serviço arrolado no documento fiscal.

**Pedido:** Representa as mercadorias ou serviços solicitados pelos clientes de um contribuinte, cujo documento provavelmente acarretará uma operação de saída.

**Item Pedido:** Identifica cada item de mercadoria ou produto pedido pelo cliente do contribuinte.

**Apelido Produto:** Corresponde ao nome abreviado que o produto recebe no estabelecimento e que normalmente aparece nos documentos fiscais.

**Tabela Produto:** Corresponde aos produtos e serviços com os quais o contribuinte trabalha, e que são de interesse do fisco para efeito de auditoria fiscal.

**Código Fiscal Operação:** Corresponde ao tipo de operação efetuada com mercadorias pelos contribuintes.

**Grupo Código Fiscal Operação:** Corresponde a uma tabela que serve para agrupar os tipos de operações efetuadas com mercadorias, pelos contribuintes.

**SubGrupo Código Fiscal Operação:** Corresponde a uma tabela que serve para subagrupar os tipos de operações efetuadas com mercadorias, pelos contribuintes.

**Fiscalização:** Representa o ato de fiscalizar, efetuado por um ou mais fiscais.

**Tipo Fiscalização:** Representa os motivos que justificam a realização da fiscalização.

**Atuação Fiscalização:** Registra o tipo de atuação de cada fiscal nas fiscalizações.

**Conta Energia Elétrica:** Corresponde aos dados sobre a conta de energia elétrica do estabelecimento.

**Titularidade Conhecida Fornecedor:** Representa a titularidade da conta de energia elétrica conhecida e informada pelo fornecedor de energia em determinado período.

**Titularidade Declarada:** Registra a titularidade declarada ao fisco estadual pelo contribuinte, referente as contas de energia elétrica por ele utilizadas.

**Produtor Agropecuário:** Entidade sub-tipo de contribuinte, que representa os produtores agropecuários, que devido as suas características tem tratamento particularizado.

**Bloco Nota Fiscal:** Corresponde aos blocos de notas fiscais distribuídos anualmente pela fazenda estadual aos produtores agropecuários, para os devidos registros.

Esta etapa é de fundamental importância para a construção de um modelo sólido e realista, e por isto recomenda-se que ela seja minuciosamente discutida e submetida a críticas rigorosas, antes de ser considerada concluída.

#### 4.2.2 DataWarehouse

Discutida, concluída e apropriadamente documentada a etapa anterior, pode-se concentrar os esforços para conhecer o contexto onde deverá atuar o modelo, e para estudar os sistemas existentes em seus respectivos ambientes dos dados.

O contexto engloba, além dos requisitos da Secretaria da Fazenda, a infraestrutura de hardware, de software, de pessoal, de telecomunicação e de instalações físicas, enquanto

o estudo dos sistemas existentes deve ser concentrado principalmente na análise dos dados que se encontram distribuídos pelas diversas bases que são utilizadas.

Usualmente estas bases estão concentradas numa única sede, e seus dados são utilizados por vários sistemas desenvolvidos internamente, para atender a objetivos específicos.

Para exemplificar, vai ser apresentado um destes sistemas, básico para todas as SEFs, onde as diferenças ficam por conta das plataformas de hardware e ferramentas de software que são utilizadas, bem como do volume de dados gerenciados.

### **Sistema de Cadastro de Contribuintes do ICMS**

- **Objetivo**

O sistema contém as informações indispensáveis à identificação, localização e classificação dos contribuintes que efetuam operações relativas à circulação de mercadorias em trânsito (compras e vendas) e a prestação de serviços.

- **Plataforma**

Hardware: Processador IBM 9672-R63, 120 Mips, 1Gb de Memória Central, Controladora de Comunicação CPM 1455-02, Impressora Laser XEROX X-4090, entrada de dados em Terminais IBM 3270 ou similar e microcomputador simulando terminal.

Software: Sistema Operacional MVS/XA 2.2.3, Linguagens ADS/COBOL/EASYTRIEVE, Banco de Dados IDMS 12.0, Gerenciador de transações On-Line IDMS/DC 12.0

Existem também diversos sistemas *stand alone* auxiliares, que utilizam hoje os recursos tecnológicos que eram disponíveis nos períodos ao longo dos quais os mesmos foram desenvolvidos.

Estes sistemas tornam árdua a tarefa de obtenção da informação desejada, exigindo dos usuários esforços desnecessários e muitas vezes acima das suas possibilidades.

Adicionalmente, após este processo multi-tecnológico de obtenção de informações, os usuários ainda precisam transportar dados para planilhas de cálculo e processadores de texto, para fazerem tabulações, análises complementares e apresentações finais dos relatórios, que se mostram pouco confiáveis, inconsistentes entre si e ineficazes para o combate a sonegação.

Devido a estes fatos, alguns problemas críticos se configuraram, podendo ser citados entre eles, a falta de integração entre os sistemas logicamente integráveis, e principalmente os problemas relacionados aos dados, tais como:

- Duplicidade – múltiplos cadastros contendo os mesmos atributos de dados.
- Inconsistência – entidades contendo atributos conflitantes.
- Falta de Integridade – referência a fatos inexistentes.
- Não confiabilidade – recuperação de dados sensível à fonte utilizada.
- Padronização não única – utilização de convenções diferentes para representar o mesmo atributo.

A partir deste ponto fica identificada com clareza a necessidade de se projetar e construir um *datawarehouse* para poder atender aos requisitos específicos da Secretaria da Fazenda.

#### **4.2.2.1 Como deve ser idealizado**

Uma etapa que não pode ser negligenciada é a análise da documentação dos sistemas existentes, quando disponíveis, análises estas que devem ser complementadas com a realização de eficazes reuniões, previamente planejadas.

Nestas reuniões devem ser ouvidas e questionadas pessoas envolvidas com o processo de tributação, nos níveis estratégico, tático e operacional.

Todo o esforço deve ser feito para não deixar de considerar elementos importantes para o perfeito entendimento do contexto, e nesta linha, devem ser consultados Tomadores de Decisão, Gerentes, Delegados Regionais, Auditores Fiscais, Fiscais de Campo, Analistas de Sistemas, Contadores Externos e o Conselho de Contribuintes.

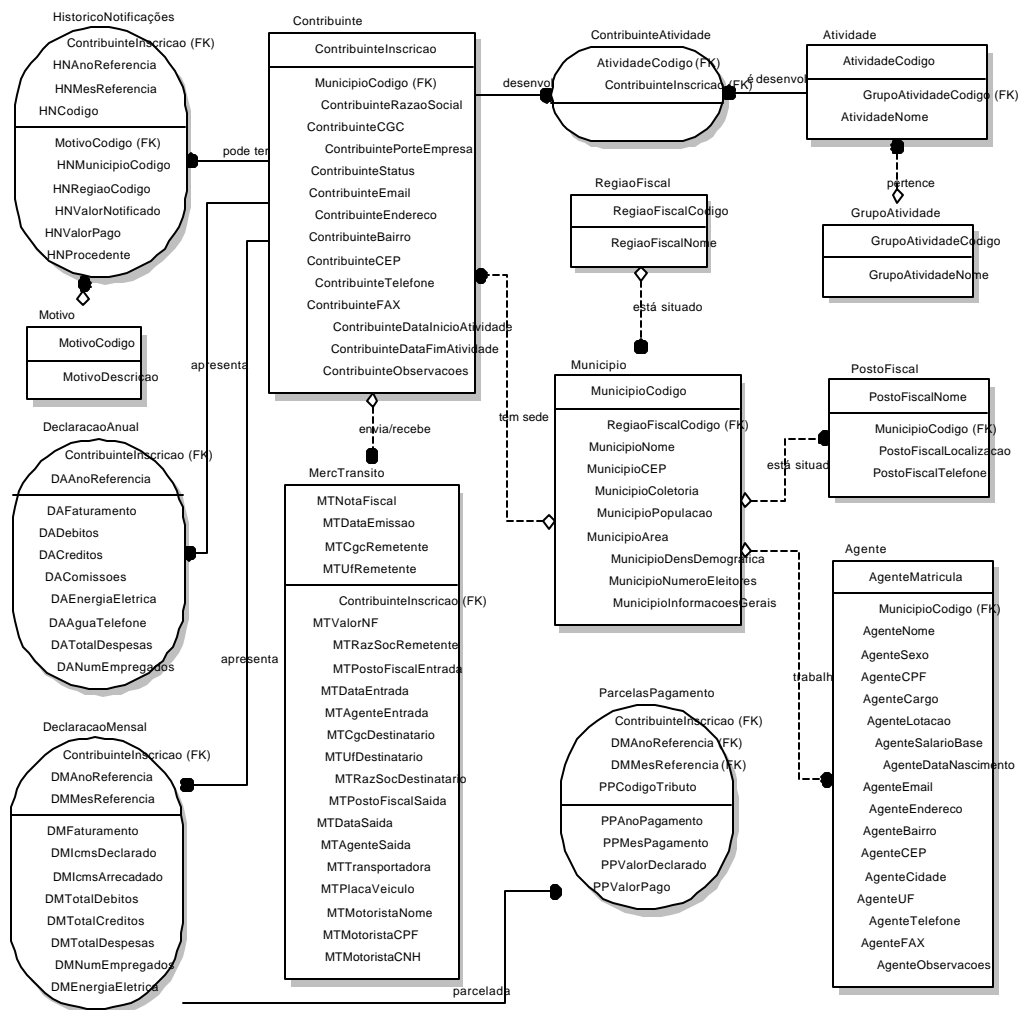
Entre os Tomadores de Decisão devem ser incluídos o Governador do Estado e o Secretário Estadual da Fazenda.

Uma vez consciente que estas recomendações foram seguidas, pode-se partir para o projeto do *datawarehouse* específico que se irá construir.

Identificam-se então nesta ordem, as Entidades, Atributos e Relacionamentos pertinentes a área específica do *datawarehouse*, em paralelo com a construção do Dicionário de Dados, visando uniformizar a nomenclatura, definição, significado e tipo de dados que se vai utilizar no modelo.

Devido aos requisitos de funcionamento previamente estabelecidos, necessário se faz um nível dual de granularidade: o primeiro, chamado aqui de Modelo de Dados de Trabalho, de baixa granularidade, sobre o qual se deverá fazer um pré-processamento visando a geração de sínteses de natureza geográfica, temporal e de tipo de atividade econômica, e o segundo, denominado aqui de Modelo de Dados Teórico, de alta granularidade, que será povoado com o resultado do pré-processamento e sobre o qual serão desenvolvidos os algoritmos do *datamining*.

Um exemplo do Modelo de Dados Teórico do *datawarehouse* proposto, passa a ser apresentado em seguida:





| Entidade CONTRIBUINTE - Pessoa física ou jurídica que contribui para a arrecadação estadual |   |      |      |          |
|---|---|------|------|----------|
| Nome do Atributo  | Descrição do Atributo                         | Tipo | Tam. | Chave    |
| ContribuinteInscricao   | Inscrição estadual do contribuinte            | Alfa | 14   | Sim      |
| MunicipioCodigo   | Código do município                           | Alfa | 05   | Estrang. |
| ContribuinteRazaoSocial   | Razão social do contribuinte                  | Alfa | 60   |          |
| ContribuinteNomeFantasia  | Nome fantasia do contribuinte                 | Alfa | 40   |          |
| ContribuinteCGC   | CGC do contribuinte                           | Alfa | 14   |          |
| ContribuintePorteEmpresa  | Porte da empresa                              | Alfa | 01   |          |
| ContribuinteStatus  | Situação cadastral do contribuinte            | Alfa | 20   |          |
| ContribuinteEmail   | Email do contribuinte                         | Alfa | 30   |          |
| ContribuinteEndereco  | Endereço do contribuinte                      | Alfa | 50   |          |
| ContribuinteBairro  | Bairro do contribuinte                        | Alfa | 25   |          |
| ContribuinteCEP   | CEP do contribuinte                           | Alfa | 08   |          |
| ContribuinteTelefone  | Telefone do contribuinte                      | Alfa | 14   |          |
| ContribuinteFAX   | FAX do contribuinte                           | Alfa | 14   |          |
| ContribuinteDataInicAtiv  | Data de início das atividades do contribuinte | Data | --   |          |
| ContribuinteDataFimAtiv   | Data de fim das atividades do contribuinte    | Data | --   |          |
| ContribuinteObservacoes   | Observações sobre o contribuinte              | Alfa | 255  |          |

Esta é a entidade básica do Modelo de Dados Teórico utilizado no *datawarehouse*, e tem como chave primária a Inscrição estadual do contribuinte, que é fornecido pela Junta Comercial, quando do registro de cada novo contribuinte. Este código é composto de até quatorze caracteres e é único, ou seja, não podem existir dois ou mais contribuintes que possuam a mesma inscrição.

O código do município serve para propósitos de filtragem geográfica dos contribuintes, enquanto a razão social e o nome fantasia propiciam importantes maneiras secundárias, alternativas e incrementais para identificação de empresas. O código do município possui integridade referencial, que é verificada na entidade Município. O CGC é usado com propósitos de validação e verificação de consistência. O porte da empresa é um sinalizador que informa o tipo de obrigação tributária a que está sujeita a empresa em

análise. O status indica a situação cadastral da empresa (suspensão de ofício, baixa voluntária, baixa de ofício, alteração, recadastramento, etc). As datas de início e fim de atividade auxiliam na otimização da recuperação de informações.

Os demais atributos são usados para facilitar o acesso dos fiscais aos locais das suas visitas.

| Entidade GRUPOATIVIDADE – Agrupamento por afinidade, de atividades econômicas exercidas pelos contribuintes. |  |      |      |       |
|--|--|------|------|-------|
| Nome do Atributo   | Descrição do Atributo                    | Tipo | Tam. | Chave |
| GrupoAtividadeCodigo   | Código do grupo de atividades econômicas | Alfa | 02   | Sim   |
| GrupoAtividadeNome   | Nome do grupo de atividades econômicas   | Alfa | 20   |       |

Esta entidade tem por finalidade representar os grupos de atividades de interesse econômico para o Estado, como por exemplo, Indústria, Comércio Atacado, Comércio a Varejo, Prestação de Serviços, Produção Rural, Extração Mineral, etc.

Tem como chave primária o código do grupo de atividades econômicas, que está associado univocamente ao outro atributo, que representa o nome do grupo de atividades econômicas.

| Entidade ATIVIDADE - Representa as atividades econômicas exercidas pelos contribuintes. |  |      |      |          |
|---|--|------|------|----------|
| Nome do Atributo  | Descrição do Atributo                    | Tipo | Tam. | Chave    |
| AtividadeCodigo   | Código da atividade econômica            | Alfa | 05   | Sim      |
| GrupoAtividadeCodigo  | Código do grupo de atividades econômicas | Alfa | 02   | Estrang. |
| AtividadeNome   | Nome da atividade econômica              | Alfa | 60   |          |

Esta entidade tem por finalidade conter os dados das atividades propriamente ditas. O modelo engloba e trabalha com todas as atividades econômicas de interesse do Estado, cada uma das quais possui um código próprio, utilizado pelo modelo como chave primária.

A integridade referencial do Código do Sub Grupo de Atividades está incluída no *datawarehouse*, bem como os dois índices secundários criados para minimização do tempo de recuperação de informações, denominados PorSubGrupo e PorNome .

| Entidade REGIAOFISCAL – Representa as unidades setoriais de fiscalização da Secretaria da Fazenda. |                         |      |      |       |
|--|-------------------------|------|------|-------|
| Nome do Atributo   | Descrição do Atributo   | Tipo | Tam. | Chave |
| RegiaoFiscalCodigo   | Código da região fiscal | Alfa | 02   | Sim   |
| RegiaoFiscalNome   | Nome da região fiscal   | Alfa | 30   |       |

Esta entidade armazena as regiões fiscais definidas pelo Estado em que se vai aplicar o modelo. As regiões fiscais são compostas por municípios pertencentes ao Estado, exceção feita a uma delas, que é composta por municípios pertencentes a outros estados, com os quais o Estado em foco tem relações comerciais.

O código da região fiscal é utilizado com chave primária, e o outro atributo representa o nome da região, para o qual foi definido um índice secundário de recuperação.

| Entidade MUNICIPIO - Representa as cidades do Estado. |   |          |      |          |
|---|---|----------|------|----------|
| Nome do Atributo                                      | Descrição do Atributo                         | Tipo     | Tam. | Chave    |
| MunicipioCodigo                                       | Código do município                           | Alfa     | 05   | Sim      |
| RegiaoFiscalCodigo                                    | Código da região fiscal onde está o município | Alfa     | 02   | Estrang. |
| MunicipioNome   | Nome do município                             | Alfa     | 30   |          |
| MunicipioCEP  | CEP do município                              | Alfa     | 08   |          |
| MunicipioColetoria                                    | Identificação da coletoria                    | Alfa     | 40   |          |
| MunicipioPopulacao                                    | População do município                        | Numérico | --   |          |
| MunicipioArea   | Área do município                             | Numérico | --   |          |
| MunicipioDensDemog                                    | Densidade demográfica do município            | Numérico | --   |          |
| MunicipioNumEleitores                                 | Número de eleitores do município              | Numérico | --   |          |
| MunicipioInformGerais                                 | Informações gerais sobre o município          | Memo     | --   |          |

Esta entidade armazena dados sobre os municípios pertencentes às regiões cadastradas na entidade Região. Pode conter municípios pertencentes a outros Estados, além é claro, de todos os municípios do estado em consideração.

Vale ressaltar que, por ter o modelo uma arquitetura aberta, novos atributos podem ser incorporados a ele, e um exemplo disto foi a inclusão dos atributos população, área, densidade demográfica e número de eleitores dos municípios, que inicialmente não faziam parte do modelo. O código do município (padrão já existente), é utilizado como chave primária .

A integridade referencial do código da região fiscal está incluída no *datawarehouse*, bem como os dois índices secundários criados para minimização do tempo de recuperação de informações (PorNomeMunicipio e PorRegiao) .

| Entidade CONTRIBUINTEATIVIDADE – Representa as atividades econômicas exercidas pelo contribuinte. |  |      |      |       |
|---|--|------|------|-------|
| Nome do Atributo  | Descrição do Atributo                  | Tipo | Tam. | Chave |
| AtividadeCodigo   | Código da atividade econômica exercida | Alfa | 05   | Sim   |
| ContribuinteInscricao   | Inscrição estadual do contribuinte     | Alfa | 14   | Sim   |

Esta entidade relaciona o contribuinte com as atividades econômicas exercidas por ele.

A chave primária é composta pela concatenação dos atributos código da atividade econômica com inscrição estadual, para os quais são verificadas as suas integridades referenciais nas entidades Contribuinte e Atividade, respectivamente.

| Entidade POSTOFISCAL - Representação da SEF com o intuito de efetuar, em caráter permanente, atividades de fiscalização de mercadorias em trânsito. |  |      |      |          |
|---|--|------|------|----------|
| Nome do Atributo  | Descrição do Atributo                        | Tipo | Tam. | Chave    |
| PostoFiscalNome   | Nome do posto fiscal                         | Alfa | 25   | Sim      |
| MunicipioCodigo   | Código do município onde está o posto fiscal | Alfa | 05   | Estrang. |
| PostoFiscalLocalização  | Descrição da localização do posto fiscal     | Alfa | 80   |          |

|                     |                          |      |    |  |
|---------------------|--------------------------|------|----|--|
| PostoFiscalTelefone | Telefone do posto fiscal | Alfa | 14 |  |
|---------------------|--------------------------|------|----|--|

Esta entidade armazena dados sobre os postos fiscais sob a responsabilidade da SEF.

Tem como chave primária o nome do posto fiscal e o código do município tem a sua integridade referencial verificada na entidade Município.

Entidade **AGENTE** – Funcionário capacitado e autorizado a efetuar oficialmente atividades de fiscalização junto ao contribuinte.

| Nome do Atributo     | Descrição do Atributo                         | Tipo  | Tam. | Chave    |
|----------------------|---|-------|------|----------|
| AgenteMatricula      | Matricula do agente                           | Alfa  | 05   | Sim      |
| MunicipioCodigo      | Código do município onde está lotado o agente | Alfa  | 05   | Estrang. |
| AgenteNome           | Nome do agente                                | Alfa  | 40   |          |
| AgenteSexo           | Sexo do agente                                | Alfa  | 01   |          |
| AgenteCPF            | CPF do agente                                 | Alfa  | 14   |          |
| AgenteCargo          | Cargo do agente                               | Alfa  | 25   |          |
| AgenteLotacao        | Lotação do agente                             | Alfa  | 25   |          |
| AgenteSalarioBase    | Salário base do agente                        | Money | --   |          |
| AgenteDataNascimento | Data de nascimento do agente                  | Data  | --   |          |
| AgenteEmail          | Email do agente                               | Alfa  | 30   |          |
| AgenteEndereco       | Endereço do agente                            | Alfa  | 50   |          |
| AgenteBairro         | Bairro do agente                              | Alfa  | 25   |          |
| AgenteCEP            | CEP do agente                                 | Alfa  | 08   |          |
| AgenteCidade         | Cidade do agente                              | Alfa  | 25   |          |
| AgenteUF             | UF do agente                                  | Alfa  | 02   |          |
| AgenteTelefone       | Telefone do agente                            | Alfa  | 14   |          |
| AgenteFAX            | FAX do agente                                 | Alfa  | 14   |          |
| AgenteObservacoes    | Observações sobre do agente                   | Memo  | --   |          |

Esta entidade armazena dados de identificação e localização dos agentes fiscais. Tem como chave primária a matrícula do agente, e o código do município tem a sua integridade referencial verificada na entidade Município.

| Entidade DECLARACAOMENSAL – conjunto mensal de informações serem fornecidas pelo contribuinte |   |         |      |       |
|---|---|---------|------|-------|
| Nome do Atributo  | Descrição do Atributo                       | Tipo    | Tam. | Chave |
| ContribuinteInscricao   | Inscrição estadual do contribuinte          | Alfa    | 25   | Sim   |
| DMAnoReferencia   | Ano de referência da declaração             | Alfa    | 04   | Sim   |
| DMMesReferencia   | Mês de referência da declaração             | Alfa    | 02   | Sim   |
| DMFaturamento   | Faturamento do contribuinte no mês          | Money   |      |       |
| DMIcmsDeclarado   | Icms declarado pelo contribuinte no mês     | Money   |      |       |
| DMIcmsArrecadado  | Icms arrecadado do contribuinte no mês      | Money   |      |       |
| DMTotalDebitos  | Total dos débitos do contribuinte no mês    | Money   |      |       |
| DMTotalCreditos   | Total dos créditos do contribuinte no mês   | Money   |      |       |
| DMTotalDespesas   | Total das despesas do contribuinte no mês   | Money   |      |       |
| DMNumEmpregados   | Número de empregados do contribuinte no mês | Inteiro |      |       |

Esta entidade armazena os dados mensais de faturamento, icms declarado, icms arrecadado, total de débitos, total de créditos, total das despesas e número de empregados, para cada contribuinte oficialmente registrado no Estado.

A chave primária é composta pela concatenação dos atributos inscrição estadual do contribuinte com ano de referência e com mês de referência. A integridade referencial da inscrição estadual do contribuinte está incluída no *datawarehouse*.

| Entidade DECLARACAOANUAL – conjunto anual de informações a serem fornecidas pelo contribuinte |   |         |      |       |
|---|---|---------|------|-------|
| Nome do Atributo  | Descrição do Atributo                           | Tipo    | Tam. | Chave |
| ContribuinteInscricao   | Inscrição estadual do contribuinte              | Alfa    | 25   | Sim   |
| DAAnoReferencia   | Ano de referência da declaração (exercício)     | Alfa    | 04   | Sim   |
| DAFaturamento   | Faturamento total do contribuinte no exercício  | Money   |      |       |
| DATotalDebitos  | Total dos débitos do contribuinte no exercício  | Money   |      |       |
| DATotalCreditos   | Total dos créditos do contribuinte no exercício | Money   |      |       |
| DAComissoes   | Total das comissões pagas no exercício          | Money   |      |       |
| DAEnergiaElétrica   | Total anual das despesas com energia elétrica   | Money   |      |       |
| DATelefone  | Total anual das despesas com telefone           | Money   |      |       |
| DATotalDespesas   | Total das despesas do contribuinte exercício    | Money   |      |       |
| DANumEmpregados   | Número de empregados no exercício               | Inteiro |      |       |

Esta entidade armazena os dados anuais de faturamento, débitos no exercício, créditos no exercício, comissões, conta de telefone, conta de energia elétrica, total das despesas e número de empregados, para cada contribuinte oficialmente registrado no Estado.

A chave primária é composta pela concatenação dos atributos inscrição estadual do contribuinte com ano de referência. A integridade referencial do Código do contribuinte está incluída no *datawarehouse*.

| Entidade PARCELAPAGAMENTO – representa os valores dos tributos que estão sendo pagos de forma parcelada |                                    |       |      |       |
|---|------------------------------------|-------|------|-------|
| Nome do Atributo  | Descrição do Atributo              | Tipo  | Tam. | Chave |
| ContribuinteInscricao   | Inscrição estadual do contribuinte | Alfa  | 25   | Sim   |
| PPAnoReferencia   | Ano de referência da declaração    | Alfa  | 04   | Sim   |
| PPMesReferencia   | Mês de referência da declaração    | Alfa  | 02   | Sim   |
| PPTributoCodigo   | Código do tributo                  | Alfa  | 05   | Sim   |
| PPAnoPagamento  | Ano do pagamento da parcela        | Alfa  | 04   |       |
| PPMesPagamento  | Mês do pagamento da parcela        | Alfa  | 02   |       |
| PPValorDeclarado  | Valor da parcela declarado         | Money | --   |       |
| PPValorPago   | Valor da parcela pago              | Money | --   |       |

Esta entidade tem por objetivo conter os dados do parcelamento de tributos pagos mensalmente, por cada contribuinte oficialmente registrado no Estado.

Os dados envolvidos são: inscrição estadual do contribuinte, ano de referência da declaração, mês de referência, código do tributo, ano do pagamento do tributo, mês do pagamento do tributo, valor declarado e valor pago.

A chave primária é composta pela concatenação dos atributos inscrição estadual do contribuinte com ano de referência da declaração, com mês de referência da declaração e com o código do tributo. A integridade referencial da inscrição estadual do contribuinte está incluída no *datawarehouse*.

| Entidade MOTIVO – representa os motivos pelos quais o contribuinte pode ser notificado |                                 |         |      |       |
|--|---------------------------------|---------|------|-------|
| Nome do Atributo   | Descrição do Atributo           | Tipo    | Tam. | Chave |
| MotivoCodigo   | Código do motivo da notificação | Inteiro | --   | Sim   |
| MotivoDescricao  | Motivo propriamente dito        | Alfa    | 80   |       |

Esta entidade armazena os eventos passíveis de notificação, ou seja, quando um contribuinte recebe uma notificação, junto com ela é apresentado o motivo da mesma, que é obtido analisando-se o regulamento do ICMS.

O código do motivo da notificação é um número inteiro, utilizado como chave primária, e o motivo é uma descrição da razão da notificação, utilizando até 80 caracteres.

| Entidade HISTORICONOTIFICACAO – contém todas as notificações efetuadas pela Secretaria da Fazenda |  |        |      |          |
|---|--|--------|------|----------|
| Nome do Atributo  | Descrição do Atributo                    | Tipo   | Tam. | Chave    |
| ContribuinteInscricao   | Inscrição estadual do contribuinte       | Alfa   | 14   | Sim      |
| HNAnoReferencia   | Ano de referência da notificação         | Alfa   | 04   | Sim      |
| HNMesReferencia   | Mês de referência da notificação         | Alfa   | 02   | Sim      |
| HNNotificacaoCodigo   | Código da notificação                    | Alfa   | 10   | Sim      |
| MotivoCodigo  | Código do motivo da notificação          | Alfa   | 03   | Estrang. |
| HNMunicipioCodigo   | Código do município do contribuinte      | Alfa   | 05   |          |
| HNRegiaoCodigo  | Código da região fiscal do contribuinte  | Alfa   | 02   |          |
| HNValorNotificado   | Valor da notificação                     |        | --   |          |
| HNValorPago   | Valor pago                               |        | --   |          |
| HNProcedente  | Indicativo de procedência da notificação | Lógico | --   |          |

Esta entidade contém o histórico das notificações por contribuinte. Tem como chave primária a concatenação dos atributos inscrição estadual do contribuinte com ano de referência da notificação, com mês de referência da notificação e com o código da notificação. Possui ainda os seguintes atributos: código do motivo, com integridade referencial verificada na entidade Motivo; valor da notificação, que representa o valor da multa a que está sujeito o contribuinte que foi notificado por alguma irregularidade, cuja explicação é obtida através da pesquisa na entidade anterior, utilizando como chave



o código do motivo especificado nesta entidade; valor pago, que registra o valor da multa que efetivamente foi paga pelo contribuinte notificado; e procedente, *flag* indicando verdadeiro ou falso através de T ou F respectivamente. Os atributos código do município e código da região fiscal do contribuinte foram inseridos nesta tabela visando a melhoria de performance dos algoritmos de pesquisa que serão utilizados.

A integridade referencial da inscrição estadual do contribuinte está incluída no *datawarehouse*.

| Entidade <b>MERCTRANSITO</b> – Informações sobre as mercadorias que transitam pelo estado, independente da origem ou destino das mesmas. |   |       |      |          |
|--|---|-------|------|----------|
| Nome do Atributo   | Descrição do Atributo   | Tipo  | Tam. | Chave    |
| MTNotaFiscal   | Identificação do nota fiscal  | Alfa  | 20   | Sim      |
| MTDataEmissao  | Data de emissão da nota fiscal  | Data  | --   | Sim      |
| MTCgcRemetente   | Cgc do remetente da mercadoria  | Alfa  | 14   | Sim      |
| MTUfRemetente  | Uf do remetente da mercadoria   | Alfa  | 02   | Sim      |
| ContribuinteInscricao  | Inscrição estadual do contribuinte                                    | Alfa  | 14   | Estrang. |
| MTValorNF  | Valor da Nota Fiscal  | Money | --   |          |
| MTRazSocRemetente  | Razão social do remetente da mercadoria                               | Alfa  | 60   |          |
| MTPostoFiscalEntrada   | Posto fiscal por onde a mercadoria entrou no estado                   | Alfa  | 10   |          |
| MTDataEntrada  | Data de passagem pelo posto fiscal de entrada                         | Data  | --   |          |
| MTAgenteEntrada  | Matrícula do agente que registrou a operação de entrada da mercadoria | Alfa  | 10   |          |
| MTCgcDestinatário  | Cgc do destinatário da mercadoria                                     | Alfa  | 14   |          |
| MTUfDestinatário   | Uf do destinatário da mercadoria                                      | Alfa  | 02   |          |
| MTRazSocDestinatario   | Razão social do destinatário da mercadoria                            | Alfa  | 60   |          |
| MTPostoFiscalSaida   | Posto fiscal por onde saiu a mercadoria                               | Alfa  | 10   |          |
| MTDataSaida  | Data de passagem da mercadoria no posto fiscal de saída               | Data  | --   |          |
| MTAgenteSaida  | Matrícula do agente que registrou a operação de saída da mercadoria   | Alfa  | 10   |          |
| MTTransportadora   | Nome da empresa transportadora  | Alfa  | 40   |          |
| MTPlacaVeiculo   | Placa do veículo que transportou a mercadoria                         |       | 08   |          |
| MTMotoristaNome  | Nome do motorista do veículo que fez o transporte da mercadoria       |       | 40   |          |

|                |   |  |    |  |
|----------------|---|--|----|--|
|                | transporte da mercadoria  |  |    |  |
| MTMotoristaCPF | Cpf do motorista do veículo que fez o transporte da mercadoria                              |  | 11 |  |
| MTMotoristaCNH | Carteira nacional de habilitação do motorista do veículo que fez o transporte da mercadoria |  | 12 |  |

Esta entidade contém o registro diário sobre as mercadorias que transitam pelo estado, independente da origem ou destino das mesmas. Tem como chave primária a concatenação dos atributos número da nota fiscal, com data de emissão, cgc e uf do remetente. Os demais atributos são utilizados para grupamentos, cálculos, fiscalizações e auditorias.

#### 4.2.2.2 Migração dos Dados

Após a conclusão do projeto lógico e a implementação do projeto físico do *datawarehouse*, chega a hora da migração dos dados.

A Figura 8, apresentada a seguir, ajuda a entender como é o esquema:

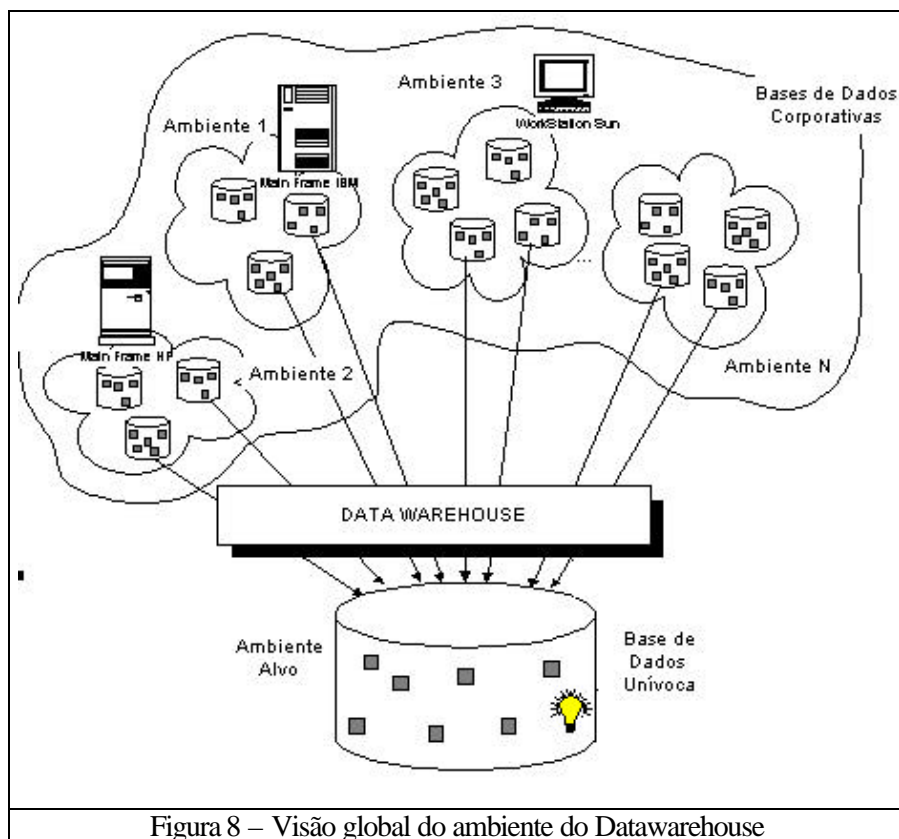


Figura 8 – Visão global do ambiente do Datawarehouse

Invariavelmente vai ser encontrada uma diversidade de ambientes de hardware e software onde o modelo será aplicado, ambientes estes que foram herdados ao longo da história da Organização.

Estes ambientes gerenciam suas próprias bases de dados e contém replicação de tabelas, de dados em tabelas e multiplicidade de processos para obtenção de consultas e relatórios.

Dependendo destes fatores, é possível se obter resultados diferentes para a mesma solicitação.

Convém ressaltar que estes problemas, não são características exclusivas das SEFs, pelo contrário, eles são comuns à grande maioria das organizações que trabalham com processamento eletrônico de dados, principalmente aquelas que já fazem isto há muito

tempo, pois elas passaram por várias fases de mudanças tecnológicas e sofreram todo tipo de pressão para fazer com que os sistemas fossem desenvolvidos de qualquer maneira; o que importava era que os sistemas ficassem prontos e entrassem logo em produção.

Como normalmente existe uma inércia nas organizações, foi possível trabalhar até hoje desta forma. Atualmente isto não é mais viável, e a razão é muito simples: Sobrevivência da organização. Vivemos na sociedade da informação, e as empresas que não puderem dispor destas informações com confiabilidade, flexibilidade e rapidez, seguramente perderão boas oportunidades de negócios e ficarão à mercê do fracasso.

A fase de migração envolve um processamento pesado, onde precisam ser desenvolvidos vários programas *stand alones* interativos, com o objetivo de pesquisar os dados (vide Figura 8), eliminar inconsistências, filtrá-los, transformá-los num padrão único, fazer a conversão de tecnologias, efetuar um pré-processamento e povoar o *datawarehouse*, de acordo com o modelo que foi projetado.

Esta fase deve ser desenvolvida gradual e interativamente, analisando cuidadosamente os resultados que serão produzidos por cada um dos processamentos realizados.

A cada nova descoberta, novas alterações se tornarão necessárias e deverão ser feitas, até se chegar a configuração final apresentada no modelo.

Cuidados especiais deverão ser tomados no povoamento das entidades, pois, devido as definições de integridades referenciais, que foram utilizadas nos casos onde eram imprescindíveis, a migração precisará obedecer uma sequência preestabelecida, isto é, primeiro se povoam as entidades independentes, procurando grupá-las de acordo com suas afinidades, para em seguida ir povoando as entidades em função dos seus graus de dependência.

Para citar um exemplo, a entidade Contribuinte, que é básica para o funcionamento do modelo, não pode ser a primeira a ser povoada, pois depende diretamente do povoamento da entidade Município, que por sua vez depende do povoamento da entidade Região.

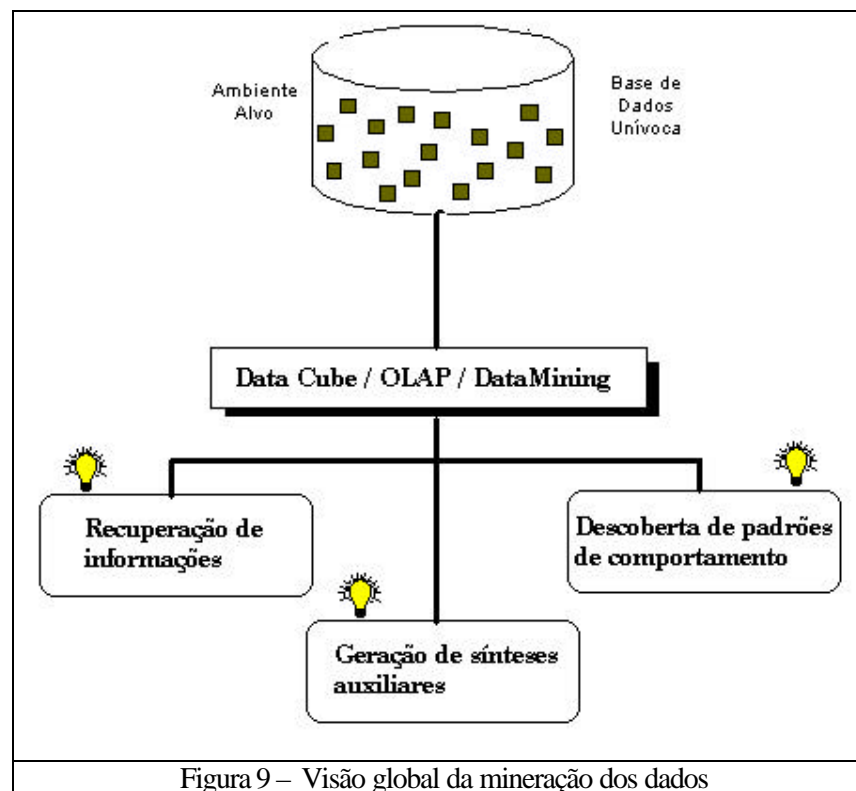
Vale ressaltar também que a quantidade de registros que será migrada para o *datawarehouse* deverá ser cuidadosamente estimada, apesar de não ser fator restritivo nem ao comportamento nem à performance deste modelo, devido à complexidade assintótica tempo dos algoritmos utilizados (BRAZ, 1980). Com relação aos equipamentos (hardware), devem ser dimensionados em função dos requisitos funcionais do sistema; exemplificando, os dispositivos de armazenamento secundário de dados, precisam ser dimensionados para suportar o volume de dados estimado para o sistema no final de um determinado período de tempo, digamos cinco anos (INMON, 1997).

#### **4.2.3 Datamining**

O modelo proposto neste trabalho, prepara o *Datawarehouse* visando o *Datamining*, que, em última instância, é quem exteriorizará as informações especificadas pela SEF em consideração. Pode-se dizer que o *Datawarehouse* é a fundação deste modelo, e deve ser construído para propiciar uma boa performance aos algoritmos do *Datamining*.

##### **4.2.3.1 Esquema utilizado**

A Figura 9, mostrada a seguir, ajuda a entender como é o esquema de atuação do *Datamining* proposto.



Conforme apresentado no Capítulo 3 – Considerações conceituais sobre *datawarehouse* e *datamining*, as organizações têm os dados, aliás, em grande quantidade, e precisam das informações envoltas neles, isto é, precisam das informações que podem ser obtidas a partir dos dados existentes, que serão utilizadas para atender cada vez melhor os seus clientes, para alocar os seus recursos de uma maneira mais eficiente, para minimizar perdas, enfim, para subsidiar a tomada de decisões.

Sabendo-se isto, foi construído o *datawarehouse* conforme apresentado nas seções anteriores deste capítulo, e vão ser apresentados agora os três grupos de funções que compõem a parte de *datamining* deste modelo.

O grupo 1, **Recuperação de informações**, foi criado para atender as necessidades de informações que são freqüentemente solicitadas. A técnica aqui utilizada pode ser

enquadrada em *OLAP-On-line Analytical Processing*, com algoritmos desenvolvidos pelo autor, que prepara as visões como num cubo de dados, propiciando um exame de todas as combinações que podem ser feitas com as informações geradas pelo modelo, ao atender uma solicitação. Neste grupo estão incluídas as seguintes consultas:

- Composição das Regiões Fiscais: Informações qualitativas e quantitativas sobre as regiões fiscais definidas pela Secretaria da Fazenda.
- Características dos Contribuintes: Informações qualitativas e quantitativas sobre os contribuintes que fazem parte do cadastro de contribuintes da Secretaria da Fazenda.
- Rotas do Estado: Trajetórias rodoviárias que interligam os municípios do Estado.
- Localização de Municípios: Identificação geográfica dos municípios do Estado.
- Pesquisa de Contribuintes: Consulta inteligente sobre a situação dos contribuintes do Estado.
- Tributos associados às atividades econômicas e períodos de tempo: Informações sobre os tributos declarados pelos contribuintes e arrecadados pelo estado.
- Mercadorias em trânsito: Informações periódicas sobre as mercadorias que transitam pelo estado, independente da origem ou destino das mesmas.

Apresentam-se agora, consulta a consulta, as informações solicitadas e a maneira de obtê-las:

### **Composição das Regiões Fiscais**

- Informações solicitadas: Nomes das regiões fiscais, Quantidade de municípios por região fiscal, Nomes dos municípios que compõem as regiões fiscais, classificados em ordem alfabética crescente, Quantidade total de municípios no estado, População das regiões fiscais, População dos municípios que compõem as regiões fiscais,

classificada em ordem numérica crescente, População total do estado, Área das regiões fiscais, Área dos municípios que compõem as regiões fiscais, classificada em ordem numérica crescente, Área total do estado, Densidade demográfica das regiões fiscais, Densidade demográfica dos municípios que compõem as regiões fiscais, classificada em ordem numérica crescente, Densidade demográfica total do estado, Quantidade de eleitores das regiões fiscais, Quantidade de eleitores dos municípios que compõem as regiões fiscais, classificada em ordem numérica crescente, Quantidade total de eleitores do estado.

❑ Algoritmo de obtenção:

Leitura do parâmetro fornecido (município, população, área, etc).

Para cada registro da entidade Região Fiscal, percorrem-se os registros da entidade Municípios, que têm a mesma chave (código da região fiscal).

Recuperam-se os dados desta entidade em função do parâmetro solicitado.

Computam-se as totalizações e faz-se o armazenamento das informações de interesse.

O algoritmo pára quando todos os registros da entidade Região Fiscal tiverem sido examinados.

❑ Forma de apresentação: Gráfica e tabela.

### **Características dos Contribuintes**

- ❑ Informações solicitadas: Nomes das regiões fiscais, Quantidade de contribuintes por região fiscal, Quantidade de contribuintes dos municípios que compõem as regiões fiscais, classificada em ordem alfabética crescente dos nomes dos municípios, Relação nominal dos contribuintes em cada um dos municípios que compõem as



regiões fiscais, juntamente com seus principais dados de identificação, Quantidade total de contribuintes no estado. Conjuntos de informações similares para as atividades econômicas exercidas pelos contribuintes do estado, ou seja, Indústria, Comércio atacado, Comércio a varejo, Prestação de serviços, Extração mineral, Produção rural e Outros. Estes conjuntos podem deverão poder ser vistos isoladamente ou em forma de combinações selecionadas livremente pelo usuário.

❑ Algoritmo de obtenção:

Leitura do parâmetro fornecido (indústria, comércio, produção rural, global, etc).

Para cada registro da entidade Contribuinte, verifica-se nas entidades Municípios e Região Fiscal a pertinência do contribuinte, usando respectivamente como chave de acesso o código do município e o código da região fiscal.

Usando agora as entidades ContribuinteAtividade, Atividade e GrupoAtividade, enquadra-se o contribuinte nas suas atividades econômicas, usando respectivamente como chave de acesso a inscrição estadual, o código da atividade econômica e o código do grupo de atividade econômica.

Fazem-se então as totalizações dos registros recuperados e o armazenamento das informações de interesse.

O algoritmo pára quando todos os registros da entidade Contribuinte tiverem sido examinados.

❑ Forma de apresentação: Gráfica e tabela.

### **Rotas do Estado**

- ❑ Informações solicitadas: Caminho a ser seguido para ir de uma localidade a outra dentro do estado.

❑ Algoritmo de obtenção:

Leitura dos pontos extremos (origem e destino) do caminho desejado.

Consulta a uma tabela auxiliar gerada por algoritmos do grupo 2, usando como chave de acesso a concatenação da origem com o destino indicado

Recuperação dos nomes e seqüências dos municípios que aparecem nesta tabela.

O algoritmo pára quando todos os registros desta tabela que têm a mesma chave tiverem sido recuperados.

❑ Forma de apresentação: Gráfica e tabela.

### **Localização de Municípios**

❑ Informações solicitadas: Posição geográfica do município no mapa do Estado.

❑ Algoritmo de obtenção:

Pesquisa do código do município, a partir do conjunto mínimo de letras do seu nome.

Consulta a uma tabela auxiliar, usando como chave de acesso o código do município, se ele foi encontrado na pesquisa. Caso contrário, o algoritmo informa e pára.

Recuperação da posição geográfica do município.

O algoritmo pára quando a posição geográfica for obtida ou não existir o código do município.

❑ Forma de apresentação: Gráfica

### **Pesquisa de Contribuintes**

- ❑ Informações solicitadas: Pertinência ou não ao cadastro de contribuintes e, caso positivo, a apresentação dos seguintes dados: inscrição estadual, cgc/cpf, razão social, nome fantasia, município onde está sediado, atividade econômica que exerce (código e descrição), endereço completo, situação cadastral, datas de início e fim de atividades.

- ❑ Algoritmo de obtenção:

Pesquisa da existência do contribuinte a partir de qualquer informação que se disponha sobre ele, como por exemplo, razão social ou parte conhecida dela, nome fantasia ou parte conhecida dela, cgc/cpf ou ainda a própria inscrição estadual, que é a chave principal de acesso.

No caso da existência do contribuinte, recuperam-se os dados supracitados a partir das entidades Contribuinte e Município, utilizando-se como chave a inscrição estadual e o código do município, respectivamente

O algoritmo pára quando forem recuperados as informações solicitadas ou após a mensagem de que o contribuinte não consta do cadastro.

- ❑ Forma de apresentação: Tabela e relatório

### **Tributos associados às atividades econômicas e períodos de tempo**

- ❑ Informações solicitadas: faturamento, icms declarado, icms arrecadado, total de débitos, total de créditos e arrecadação *per capita*, provenientes das atividades econômicas exercidas no estado, analisadas isoladamente ou em conjunto num período de tempo qualquer, a ser informado dinamicamente pelo profissional da

Secretaria da Fazenda. Estas informações devem ser grupadas também por região fiscal, por município pertencente à região fiscal e por contribuinte do município.

❑ Algoritmo de obtenção:

Leitura da(s) atividade(s) que se deseja(m) examinar.

Leitura do(s) parâmetro(s) associado(s) às atividades previamente informadas.

Recepção do intervalo de tempo a ser considerado.

Usando como básicas as entidades DeclaraçãoMensal e ParcelaPagamento, e tendo como auxiliares as entidades Contribuinte, Atividade, GrupoAtividade, RegiãoFiscal, Município e ContribuinteAtividade, enquadra-se o contribuinte nas suas atividades econômicas, usando respectivamente como chave de acesso a inscrição estadual, o código da atividade econômica e o código do grupo de atividade econômica; para o enquadramento do contribuinte no município e região fiscal, usam-se as chaves código do município e código da região fiscal, respectivamente.

Fazem-se então as totalizações dos registros recuperados e o armazenamento das informações de interesse.

O algoritmo pára quando todos os registros da entidade DeclaraçãoMensal tiverem sido examinados.

❑ Forma de apresentação: Gráfico, tabela e relatório.

**Mercadorias em trânsito:**

- ❑ Informações solicitadas: número da nota fiscal, data de emissão, valor da nota fiscal, remetente(uf, cgc, razão social), destinatário(uf, cgc, razão social), posto fiscal e data de entrada, posto fiscal e data de saída, agente no posto fiscal de entrada, agente

no posto fiscal de saída, transportadora, placa do veículo e motorista (nome, cpf e carteira nacional de habilitação).

Estas informações devem ser grupadas e sintetizadas por uf de origem, uf de destino, remetente e destinatário

❑ Algoritmo de obtenção:

Recepção da origem desejada (contribuinte, uf ou todas).

Recepção do destino desejado (contribuinte, uf ou todos).

Recepção do intervalo de tempo a ser considerado.

Usando como básica a entidade MercTransito, selecionam-se os registros que satisfazem a origem, o destino e o intervalo de tempo fornecidos.

Fazem-se então as totalizações destes registros e o armazenamento das informações de interesse.

O algoritmo pára quando todos os registros que foram selecionados na entidade MercTransito tiverem sido examinados.

❑ Forma de apresentação: Tabela e relatório.

O grupo 2, **Geração de sínteses auxiliares**, foi criado visando a melhoria de performance dos algoritmos, poupando-os da tarefa de acessar, a cada solicitação de processamento, múltiplas entidades onde se encontram os dados de interesse quantitativo, que precisam ser colocados em níveis diferenciados de síntese. A técnica aqui utilizada está fundamentada na granularidade definida no *datawarehouse*, em função da qual os algoritmos dos outros grupos aqui apresentados podem ter seus desempenhos variando no intervalo (excelente a catastrófico). Devido à característica de não volatilidade do *datawarehouse*, explicada no capítulo 3, estas sínteses são

realizadas apenas uma vez, e ocorrem automaticamente sempre que houver uma nova carga de dados.

Neste grupo estão incluídas as seguintes sínteses:

- Contribuinte Rápido: conjunto de informações que agilizam o acesso e a recuperação de dados sobre os contribuintes.
- Contribuintes Ativos Giam: conjunto dos contribuintes que têm obrigatoriedade de preencher e entregar mensalmente dados solicitados pela SEF.
- Contribuintes por Município: conjunto de informações sobre o contribuinte, sintetizado por município.
- Giam por Município: conjunto de informações sobre a Giam, sintetizado por município.
- Montagem dos perfis dos contribuintes: conjunto de informações com as características estatísticas dos dados econômico-tributários dos contribuintes.

Apresentam-se agora os dados necessários e a maneira de gerar cada uma destas sínteses:

### **Contribuinte Rápido**

- Dados necessários: inscrição estadual, código do município, código da atividade, código do grupo de atividades, razão social, cgc/cpf, código do status, código do porte, nome fantasia e código da região fiscal do contribuinte.

- Algoritmo de obtenção:

```
delete from ContribuinteRapido
```

```
insert into ContribuinteRapido
```

```
(      ContribuinteInscricao,  
      ContribuinteCodMunicipio,
```

```

ContribuinteCodAtividade,
ContribuinteGrpAtividade,
ContribuinteRazaoSocial,
ContribuinteCGC,
ContribuinteCodStatus,
ContribuinteCodPorte,
ContribuinteNomeFantasia,
ContribuinteCodRegiaoFisc )

select

c.ContribuinteInscricao,
c.ContribuinteCodMunicipio,
c.ContribuinteCodAtividade,
c.ContribuinteGrpAtividade,
c.ContribuinteRazaoSocial,
c.ContribuinteCGC,
c.ContribuinteCodStatus,
c.ContribuinteCodPorte,
c.ContribuinteNomeFantasia,
c.ContribuinteCodRegiaoFisc

from   Contribuinte c

```

- ❑ Forma de apresentação: Armazenamento em arquivo.

### **Contribuintes Ativos Giam**

- ❑ Dados necessários: inscrição estadual, código do município, código da região fiscal e razão social do contribuinte
- ❑ Algoritmo de obtenção:

```

delete from ContribuinteAtivoGiam

insert into ContribuinteAtivoGiam

{
ContribuinteInscricao,
ContribuinteCodMunicipio,
ContribuinteCodRegiaoFisc,
ContribuinteRazaoSocial }

select

c.ContribuinteInscricao,
c.ContribuinteCodMunicipio,
c.ContribuinteCodRegiaoFisc,
c.ContribuinteRazaoSocial

from   ContribuinteRapido c

where

(c.ContribuinteCodStatus = '01'
or c.ContribuinteCodStatus = '02'
or c.ContribuinteCodStatus = '03'

```

```

or c.ContribuinteCodStatus = '06'
or c.ContribuinteCodStatus = '09')
and c.ContribuinteGrpAtividade <> '07'
and c.ContribuinteGrpAtividade <> '09'
and c.ContribuinteCodAtividade <> '060000'
and c.ContribuinteCodAtividade <> '060001'
and c.ContribuinteCodAtividade <> '060002'
and c.ContribuinteCodAtividade <> '060003'
and c.ContribuinteCodAtividade <> '060004'
and c.ContribuinteCodAtividade <> '060005'
and c.ContribuinteCodAtividade <> '060006'
and c.ContribuinteCodAtividade <> '060007'
and c.ContribuinteCodAtividade <> '060008'
and c.ContribuinteCodAtividade <> '060009'
and c.ContribuinteCodAtividade <> '060010'
and c.ContribuinteCodAtividade <> '050104'

```

- ❑ Forma de apresentação: Armazenamento em arquivo.

### **Contribuintes por município**

- ❑ Dados necessários: código do município, nome do município, código da região fiscal, código do grupo de atividades e total de registros
- ❑ Algoritmo de obtenção:

```

delete from ScontMun

insert into ScontMun

( MunicipioCodigo, MunicipioNome, RegiaoFiscalCodigo,
  GrupoAtividadeCodigo, ScontMunTotal )

select
  m.MunicipioCodigo,
  m.MunicipioNome,
  m.RegiaoFiscalCodigo,
  c.ContribuinteGrpAtividade as GrupoAtividadeCodigo,
  count(*) as SContMunTotal

from   ContribuinteRapido c, Municipio m

where m.MunicipioCodigo = c.ContribuinteCodMunicipio
And (c.ContribuinteCodStatus = '01'
or   c.ContribuinteCodStatus = '02'
or   c.ContribuinteCodStatus = '03'
or   c.ContribuinteCodStatus = '06'
or   c.ContribuinteCodStatus = '09')

group by m.MunicipioCodigo,

```



```

m.MunicipioNome,
m.RegiaoFiscalCodigo,
c.ContribuinteGrpAtividade

```

- ❑ Forma de apresentação: Armazenamento em arquivo.

### Giam por Município

- ❑ Dados necessários: Inscrição estadual, código do município, código da atividade, código do grupo de atividades, razão social, cgc/cpf, código do status, código do porte, nome fantasia e código da região fiscal do contribuinte.

- ❑ Algoritmo de obtenção:

```
delete from SGiamMun
```

```
insert into SGiamMun
```

```

{
  SGiamMunAno,
  SGiamMunMes,
  SGiamMunGrpAtividade,
  SGiamMunCodMunicipio,
  RegiaoFiscalCodigo,
  SGiamMunFaturamento,
  SGiamMunValorICMSDeclarad,
  SGiamMunValorICMSPago,
  SGiamMunTotalDebitos,
  SgiamMunTotalCreditos }

```

```
select
```

```

GiamAnoReferencia      as SGiamMunAno,
GiamMesReferencia      as SGiamMunMes,
GrupoAtividadeCodigo   as SGiamMunGrpAtividade,
ContribuinteCodMunicipio as SGiamMunCodMunicipio,
RegiaoFiscalCodigo,
sum (GiamFaturamento)  as SGiamMunFaturamento,
sum (GiamValorICMSDeclarado) as SGiamMunValorICMSDeclarad,
sum (GiamValorICMSPago)  as SGiamMunValorICMSPago,
sum (GiamSaida)          as SGiamMunTotalDebitos,
sum (GiamEntrada)        as SGiamMunTotalCreditos

```

```
from
```

```

Giam
left join ContribuinteRapido
      on ContribuinteInscricao = GiamInscricao
left join Municipio

```

```

        on ContribuinteCodMunicipio = MunicipioCodigo
    left join Atividade
        on ContribuinteCodAtividade = AtividadeCodigo

    group by
        GiamAnoReferencia,
        GiamMesReferencia,
        GrupoAtividadeCodigo,
        ContribuinteCodMunicipio,
        RegiaoFiscalCodigo

```

- ❑ Forma de apresentação: Armazenamento em arquivo.

### Montagem dos perfis dos contribuintes

- ❑ Dados necessários: inscrição estadual, ano de referência, mês de referência, código da atividade, código do grupo de atividades, faixa de faturamento, código da região fiscal, código do município, e os valores de icms declarado, icms pago, entrada, retido na entrada, saída, retido na saída e faturamento.
- ❑ Algoritmo de obtenção:

```

/* Primeira parte
insert into GiamParaIndicios

(
    GiamInscricao,
    GiamMesReferencia,
    GiamAnoReferencia,

    GiamCodAtividade,
    GiamCodGrpAtividade,
    GiamCodFaixaFaturamento,
    GiamCodRegiaoFiscal,
    GiamCodMunicipio,

    GiamValorIcmsDeclarado,
    GiamValorRetidoEntrada,
    GiamValorRetidoSaida,
    GiamValorIcmsPago,
    GiamEntrada,
    GiamSaida,
    GiamFaturamento )

select
    GiamInscricao,

```

```

        GiamMesReferencia,
        GiamAnoReferencia,

        c.ContribuinteCodAtividade,
        c.ContribuinteGrpAtividade,
        ff.FaixaFaturamentoCodigo,
        c.ContribuinteCodRegiaoFisc,
        c.ContribuinteCodMunicipio,

        GiamValorIcmsDeclarado,
        GiamValorRetidoEntrada,
        GiamValorRetidoSaida,
        GiamValorIcmsPago,
        GiamEntrada,
        GiamSaida,
        GiamFaturamento

from
    Giam g,
    ContribuinteRapido c,
    FaixaFaturamento ff

where ContribuinteInscricao = GiamInscricao
    and (g.GiamFaturamento >= ff.minimo
    and g.GiamFaturamento < ff.maximo)

/* Segunda parte
insert into PerfilGiamIntermediario

(    GiamAnoReferencia,
        GiamMesReferencia,
        GiamCodAtividade,
        GiamCodFaixaFaturamento,

        MediaICMSDeclarado,
        MediaICMSPago,
        MediaRetidoEntrada,
        MediaRetidoSaida,
        MediaEntrada,
        MediaSaida,
        MediaEntradaSaida,
        TamanhoAmostra

)

select
    GiamAnoReferencia,
    GiamMesReferencia,
    GiamCodAtividade,
    GiamCodFaixaFaturamento,
    Avg(GiamValorIcmsDeclarado/GiamFaturamento)
        as MediaICMSDeclarado,
    Avg(GiamValorIcmsPago/GiamFaturamento)
        as MediaICMSPago,
    Avg(GiamValorRetidoEntrada/GiamFaturamento)
        as MediaRetidoEntrada,
    Avg(GiamValorRetidoSaida/GiamFaturamento)
        as MediaRetidoSaida,

```

```

        Avg(GiamEntrada/GiamFaturamento)
            as MediaEntrada,
        Avg(GiamSaida/GiamFaturamento)
            as MediaSaida,
        Avg(GiamEntrada/GiamSaida)
            as MediaEntradaSaida,
        count(*) as TamanhoAmostra

from
    GiamParaIndicios g

where GiamFaturamento <> 0

group by
    GiamAnoReferencia,
    GiamMesReferencia,
    GiamCodAtividade,
    GiamCodFaixaFaturamento

```

**/\* Terceira parte**

insert into PerfilGiam

```

(
    GiamAnoReferencia,
    GiamMesReferencia,
    GiamCodAtividade,
    GiamCodFaixaFaturamento,

    MediaICMSDeclarado,
    DesvioICMSDeclarado,

    MediaICMSPago,
    DesvioICMSPago,

    MediaRetidoEntrada,
    DesvioRetidoEntrada,

    MediaRetidoSaida,
    DesvioRetidoSaida,

    MediaEntrada,
    DesvioEntrada,

    MediaSaida,
    DesvioSaida,

    MediaEntradaSaida,
    DesvioEntradaSaida,

    TamanhoAmostra
)

```

select

```

    pgi.GiamAnoReferencia,
    pgi.GiamMesReferencia,
    pgi.GiamCodAtividade,
    pgi.GiamCodFaixaFaturamento,
    MediaICMSDeclarado,

```

$$\begin{aligned}
& \text{Sum}(((\text{GiamValorIcmsDeclarado}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaICMSDeclarado}) \\
& \quad * ((\text{GiamValorIcmsDeclarado}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaICMSDeclarado})) / (\text{TamanhoAmostra} - 1), \\
& \text{MediaICMSPago}, \\
& \text{Sum}(((\text{GiamValorIcmsPago}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaICMSPago}) \\
& \quad * ((\text{GiamValorIcmsPago}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaICMSPago})) / (\text{TamanhoAmostra} - 1), \\
& \text{MediaRetidoEntrada}, \\
& \text{Sum}(((\text{GiamValorRetidoEntrada}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaRetidoEntrada}) \\
& \quad * ((\text{GiamValorRetidoEntrada}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaRetidoEntrada})) / (\text{TamanhoAmostra} - 1), \\
& \text{MediaRetidoSaida}, \\
& \text{Sum}(((\text{GiamValorRetidoSaida}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaRetidoSaida}) \\
& \quad * ((\text{GiamValorRetidoSaida}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaRetidoSaida})) / (\text{TamanhoAmostra} - 1), \\
& \text{MediaEntrada}, \\
& \text{Sum}(((\text{GiamEntrada}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaEntrada}) \\
& \quad * ((\text{GiamEntrada}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaEntrada})) / (\text{TamanhoAmostra} - 1), \\
& \text{MediaSaida}, \\
& \text{Sum}(((\text{GiamSaida}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaSaida}) \\
& \quad * ((\text{GiamSaida}/\text{GiamFaturamento}) - \\
& \text{pgi.MediaSaida})) / (\text{TamanhoAmostra} - 1), \\
& \text{MediaEntradaSaida}, \\
& \text{Sum}(((\text{GiamEntrada}/\text{GiamSaida}) - \\
& \text{pgi.MediaEntradaSaida}) \\
& \quad * ((\text{GiamEntrada}/\text{GiamSaida}) - \\
& \text{pgi.MediaEntradaSaida})) / (\text{TamanhoAmostra} - 1), \\
& \text{TamanhoAmostra}
\end{aligned}$$

from

GiamParaIndicios g,  
 PerfilGiamIntermediario pgi

where

pgi.GiamAnoReferencia = g.GiamAnoReferencia  
 and pgi.GiamMesReferencia = g.GiamMesReferencia  
 and pgi.GiamCodAtividade = g.GiamCodAtividade  
 and pgi.GiamCodFaixaFaturamento = g.GiamCodFaixaFaturamento  
 and pgi.TamanhoAmostra > 6

group by

pgi.GiamAnoReferencia,  
 pgi.GiamMesReferencia,  
 pgi.GiamCodAtividade,  
 pgi.GiamCodFaixaFaturamento,  
 MediaICMSDeclarado,  
 MediaICMSPago,  
 MediaRetidoEntrada,  
 MediaRetidoSaida,  
 MediaEntrada,

MediaSaida,  
MediaEntradaSaida,  
TamanhoAmostra

- ❑ Forma de apresentação: Armazenamento em arquivo.

O grupo 3, **Descoberta de padrões de comportamento**, foi projetado para identificar os prováveis sonegadores, através da descoberta de relações entre os atributos que possam mostrar indícios evasão fiscal. Os algoritmos podem trabalhar com qualquer um dos atributos constantes do *datawarehouse*, (faturamento, icms declarado, icms pago, débitos, créditos, energia elétrica, etc), de maneira isolada ou englobando-os através de uma expressão matemática, para, no final do processamento, revelar os “Contribuintes a Serem Visitados”. Regras descobertas pelo algoritmo podem ser adicionadas à base de conhecimentos do modelo.

Neste grupo estão incluídas as seguintes consultas:

- Omissos de GIAM – Guia de Informação e Apuração Mensal
- Imposto devido declarado e não pago
- Indícios de Evasão Fiscal

Apresentam-se agora, consulta a consulta, as informações solicitadas e a maneira de obtê-las:

### **Omissos de GIAM**

- ❑ Informações solicitadas: identificação dos contribuintes que deveriam e não entregaram a Giam; as informações deverão ser apresentadas através de uma relação grupada por região fiscal/município, contendo os seguintes atributos: inscrição estadual, cgc/cpf, mês e ano da ocorrência, razão social, nome fantasia, endereço,

código do município, código do grupo de atividades, código da atividade, código do status, código do porte e código da região fiscal do contribuinte.

❑ Algoritmo de obtenção:

Leitura do intervalo de tempo que se deseja examinar( mês/ano inicial → mês/ano final).

Formação dos dois conjuntos especificados abaixo:

**Cgiam<sub>i</sub> [i:=1 to n]** → Conjunto dos contribuintes que entregaram a Giam,

**Ccontg<sub>j</sub> [j:=1 to m]** → Conjunto dos contribuintes que deveriam ter entregue a Giam,

onde **n** representa a quantidade de elementos do conjunto Cgiam, **m** representa a quantidade de elementos do conjunto Ccontg e **m** >= **n**. Observe-se aqui que

**Ccontg<sub>j</sub>** pode ser obtido diretamente do grupo 2, “Geração de sínteses auxiliares”.

Pesquisa dos elementos do conjunto Ccontg que não estão no conjunto Cgiam.

Inclusão dos elementos encontrados na relação dos contribuintes a serem visitados.

O algoritmo pára quando todos os elementos do conjunto Ccontg estiverem sido examinados.

❑ Forma de apresentação: Tabela e relatório

### **Imposto devido declarado e não pago**

- ❑ Informações solicitadas: identificação dos contribuintes que declararam mas não pagaram o imposto devido, informado na Giam; as informações deverão ser apresentadas através de uma relação grupada por região fiscal/município, contendo os seguintes atributos: inscrição estadual, cgc/cpf, mês e ano da ocorrência, razão social, nome fantasia, endereço, código do município, código do grupo de

atividades, código da atividade, código do status, código do porte e código da região fiscal do contribuinte.

❑ Algoritmo de obtenção:

Leitura do intervalo de tempo que se deseja examinar( mês/ano inicial → mês/ano final).

Formação dos dois conjuntos especificados abaixo:

**Cgiam<sub>i</sub> [i:=1 to n]** → Conjunto dos contribuintes que entregaram a GIAM,

**Cgate<sub>j</sub> [j:=1 to m]** → Conjunto dos contribuintes para o quais existem DAR/GATE,

onde **n** representa a quantidade de elementos do conjunto Cgiam, **m** representa a quantidade de elementos do conjunto Cgate e **n** >= **m**.

Pesquisa os elementos da Cgiam que estão também em Cgate e comparação dos valores declarados, informados na GIAM, com os pagos, comprovados pela GATE, em todos os elementos coincidentes obtidos pela pesquisa;

Inclusão na relação dos contribuintes a serem visitados, os contribuintes para os quais foram encontradas diferenças entre estes valores.

❑ Forma de apresentação: Tabela e relatório

### **Indícios de evasão fiscal**

- ❑ Informações solicitadas: identificação dos contribuintes que apresentam indícios de evasão fiscal; as informações deverão ser apresentadas através de uma relação grupada por região fiscal/município e atividades, contendo os seguintes atributos: inscrição estadual, cgc/cpf, razão social, nome fantasia, código e descrição do indício encontrado, mês e ano da ocorrência, endereço, código do município, código



do grupo de atividades, código da atividade, código do status, código do porte e código da região fiscal do contribuinte.

□ Algoritmo de obtenção:

Leitura do intervalo de tempo que se deseja examinar( mês/ano inicial → mês/ano final).

Formação de conjuntos, onde cada um deles engloba todos os contribuintes que atuam em cada uma das atividades econômicas do Estado. O número de conjuntos será igual ao número de atividades econômicas ativas (aquelas onde existem contribuintes).

Subdivisão destes conjuntos em função das faixas de faturamento, que podem ser informadas ou geradas por uma regra preestabelecida.

Seleção dos parâmetros para análise (faturamento, icms declarado, icms pago, débitos, créditos, energia elétrica, etc), ou expressão matemática envolvendo-os.

Mapeamento dos novos conjuntos, calculando média e desvio padrão para cada um dos parâmetros selecionados.

Vale a observação de que todo o pré-processamento necessário a este algoritmo, já foi feito no grupo 2, “Geração de sínteses auxiliares”.

Seleção para compor a relação dos contribuintes a serem visitados, daqueles contribuintes que, em qualquer uma das análises, tiveram seus posicionamentos  $k$  desvios padrões abaixo da média, onde  $k$  é um parâmetro que pode ser calibrado convenientemente. Quanto maior for o  $k$ , menor será a quantidade de contribuintes relacionados pelo algoritmo e maior a significância do resultado

□ Forma de apresentação: Tabela e relatório

## 5 RESULTADOS OBTIDOS COM O MODELO

Este capítulo apresenta uma síntese dos resultados reais, obtidos com um sistema, aqui chamado de **Inteligência Fiscal**, concebido e implementado à luz do modelo proposto, e aplicado a uma unidade da Federação Brasileira.

Por se tratar de um produto, o sistema **Inteligência Fiscal** transcende o escopo deste trabalho e, por esta razão, aqui serão feitas apenas as considerações pertinentes ao modelo.

É importante ressaltar que, anteriormente à concepção deste modelo, foi desenvolvido um protótipo para uma outra unidade da Federação Brasileira, que, utilizando dados reais de um período de vinte e dois meses, forneceu subsídios valiosos para o atual modelo.

### 5.1 Relativos ao Datawarehouse

Em função das peculiaridades da SEF onde foi implementado este modelo, (todas as SEF's tem algumas peculiaridades), o *Datawarehouse* foi projetado para ficar particularmente atento aos dados dos Contribuintes, das Mercadorias, do ICMS Declarado e do ICMS Arrecadado, que são obtidos a partir de múltiplos ambientes e bases de dados, quando do processo de migração. A seguir estão relacionados, por entidade básica, os principais problemas identificados pelo modelo.

#### 5.1.1 Contribuintes

São as pessoas físicas e jurídicas estabelecidas comercialmente no Estado em consideração, ou que lá devam recolher tributos, como por exemplo, as empresas

estabelecidas em outros estados, mas que mantém Termo de Acordo de Regimes Especiais.

A fonte primária dos dados do contribuinte é um formulário chamado Boletim de Informações Cadastrais, preenchido pelo próprio contribuinte.

Já na carga inicial do *Datawarehouse*, o modelo apontou os seguintes tipos de problemas em relação a estes dados:

➤ Identificação do Contribuinte

Nome da Empresa em branco e/ou sem sentido.

Endereço da Empresa em branco e/ou sem sentido.

Telefone inexistente e/ou sem sentido.

Inscrição Estadual duplicada.

CGC / CPF inexistente.

CGC / CPF inválido.

CGC duplicado.

CEP zerado.

➤ Localização Geográfica

Código de município inexistente.

➤ Atividade Econômica

Código de atividade inexistente.

➤ Registro

Código de registro / porte inexistente.

Datas inválidas.

Datas em formatos diferentes.

Data Final das Atividades < Data Inicial das Atividades.

➤ Participação Societária

Insuficiência de dados dos sócios.

### 5.1.2 Mercadorias

O arquivo de mercadorias (baixa granularidade) contém o registro de todas as Notas Fiscais dos produtos que passam pelos postos fiscais do Estado. Existem controles específicos para as notas originárias de outros estados e destinadas ao Estado em consideração, para as notas originárias do Estado em consideração e destinadas a outros estados, e para as notas que circulam internamente ao Estado.

Apenas para se ter uma idéia do volume de dados deste arquivo, basta saber que um dos postos fiscais de fronteira, num dos dias de grande movimento, chegou a registrar a passagem de 1.500 (um mil e quinhentos) caminhões.

Os principais erros detectados pelo modelo nos dados das mercadorias foram:

- Notas Fiscais sem número.
- Posto Fiscal de entrada inexistente.
- Posto Fiscal de saída inexistente.
- CGC do remetente inválido.
- CGC do destinatário inválido.
- Valor de Notas Fiscais fora dos limites razoáveis.
- Sigla de Estado inválida.
- Matrícula de Agentes Fiscais inválidas.
- Datas inválidas (emissão da Nota Fiscal, passagem nos Postos Fiscais de entrada/saída).
- Datas em formatos diferentes.
- Data de passagem nos Postos Fiscais < data de emissão da Nota Fiscal.

- Estado remetente = Estado destinatário # Estado em consideração.
- Placa do veículo não identificada.

### **5.1.3 ICMS Declarado**

São os dados provenientes da GIAM – Guia de Informação e Apuração Mensal, que são informados pelos contribuintes e dizem respeito à apuração do cálculo do imposto devido ao Estado.

Os principais erros detectados aqui pelo modelo foram:

- Inscrição estadual do contribuinte inexistente.
- Valor declarado fora dos limites razoáveis.
- Campos inválidos (alfa em lugar de numérico).
- Campos não preenchidos

Convém ressaltar que na ocasião, (fev/2000), estes dados estavam sendo fornecidos através de formulários em papel e através de meios magnéticos, na proporção de 33,3% e 66,7% respectivamente. Na medida em que o percentual de fornecimento destes dados diretamente em meios magnéticos for aumentado, a quantidade de erros irá diminuindo.

### **5.1.4 ICMS Arrecadado**

São os dados provenientes da GATE – Guia de Arrecadação de Tributos Estaduais, que são entregues em Bancos, de onde são transmitidos para a SEF.

Os principais erros detectados aqui pelo modelo foram:

- Inscrição estadual do contribuinte inexistente.
- Valor declarado fora dos limites razoáveis.
- Código de tributo inválido.
- Falta de correspondência entre o Icms declarado e o Icms Arrecadado

- Campos não preenchidos.

### 5.1.5 Outros Dados

O *Datawarehouse* engloba ainda algumas outras entidades, que ajudam muito o processo de análises complementares feitas sobre os dados. Entre elas de stacam-se:

- Notificações de Infração.
- Processos Administrativos Tributários.
- Dívida Ativa dos Contribuintes.

### 5.1.6 Migração

Após a análise dos erros detectados pelo Modelo, os dados autorizados pelo *DBA* – *Data Base Administrator* da SEF, foram migrados satisfatoriamente para o *Datawarehouse*. Inicialmente se havia projetado o Sistema para fazer as migrações automaticamente, a intervalos de tempo regulares, selecionados pelo Usuário. Posteriormente verificou-se a inviabilidade prática desta estratégia, pois não se teria garantia alguma de que os dados necessários para a migração já estariam disponíveis. Atualmente a migração é feita quando autorizada pelo *DBA*

Aqui valem duas observações:

- A migração é um processo demorado (a carga inicial demorou 3hs e 44min).
- Alguns erros, apesar de detectados, não podem ser corrigidos sem autorização formal do contribuinte; outros precisam da assinatura da autoridade competente.

## 5.2 Relativos ao Datamining

Uma das características deste modelo, que a experiência prática mostrou ter sido bastante apropriada, foi a preparação do *Datawarehouse* visando o *Datamining*.

O *Datawarehouse* é a fundação deste modelo, e foi construído para propiciar uma boa performance aos algoritmos do *Datamining*.

Assim, durante a fase de Migração foram geradas todas as sínteses necessárias à minimização do tempo de acesso ao dispositivo de armazenamento.

Desta forma conseguiu-se obter uma excelente performance dos algoritmos de *Datamining*.

Uma grande parte do trabalho pesado de preparação de entidades com diferentes níveis de granularidade é feita através de pré-processamentos, que ficam a cargo dos algoritmos do *Datawarehouse*.

A seguir estão relacionadas, em três grupos, os benefícios obtidos pela SEF utilizando este modelo:

#### **5.2.1 Aspectos Contextuais**

Aqui caracterizados como informações de infraestrutura, que são geradas pelo modelo, visando subsidiar os Usuários nos aspectos relativos à confiabilidade dos dados com os quais está trabalhando, e baseados nos quais as decisões serão tomadas.

Isto foi muitíssimo importante para dar credibilidade ao Sistema.

Enquadram-se neste grupo, os dados que possuem baixa rotatividade e que já se encontram devidamente corrigidos e refinados.

As informações, sempre que possível, foram apresentadas em forma de gráficos, visando atender aos anseios dos Usuários e ajudá-los a observar fatos relevantes.

Seguem alguns exemplos de informações que foram geradas e pertencem a este grupo:

### 5.2.1.1 Composição das Regiões Fiscais

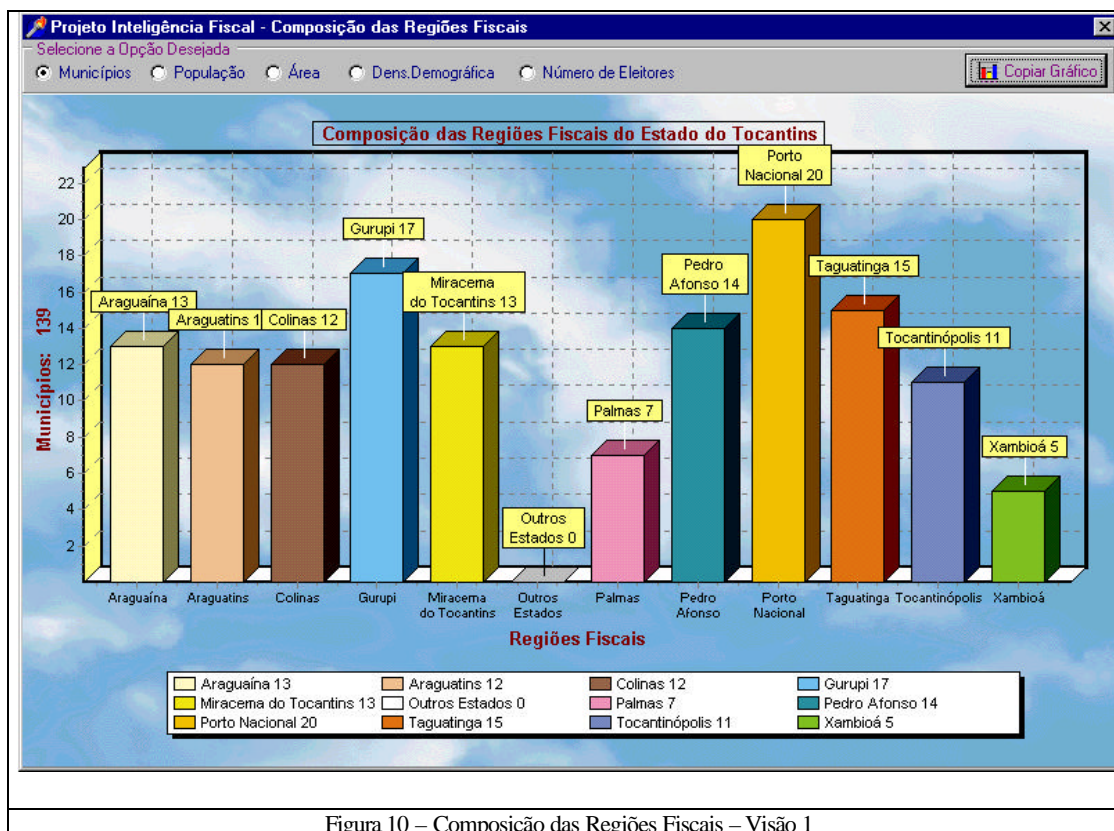


Figura 10 – Composição das Regiões Fiscais – Visão 1

Observando-se a figura acima, pode-se ver que, de acordo com a opção selecionada, foram apresentadas as quantidades de Municípios existentes em cada uma das regiões fiscais do Estado. O eixo y (vertical) mostra a quantidade total de municípios no Estado. Nesta mesma consulta, poderia ser apresentada a População, a Área, a Densidade Demográfica, ou o Número de Eleitores.

Ainda nesta mesma consulta pode-se obter outro nível de detalhamento, para mostrar os nomes dos **n** Municípios da região selecionada, que aparecem classificados em ordem alfabética crescente (Figura 11).



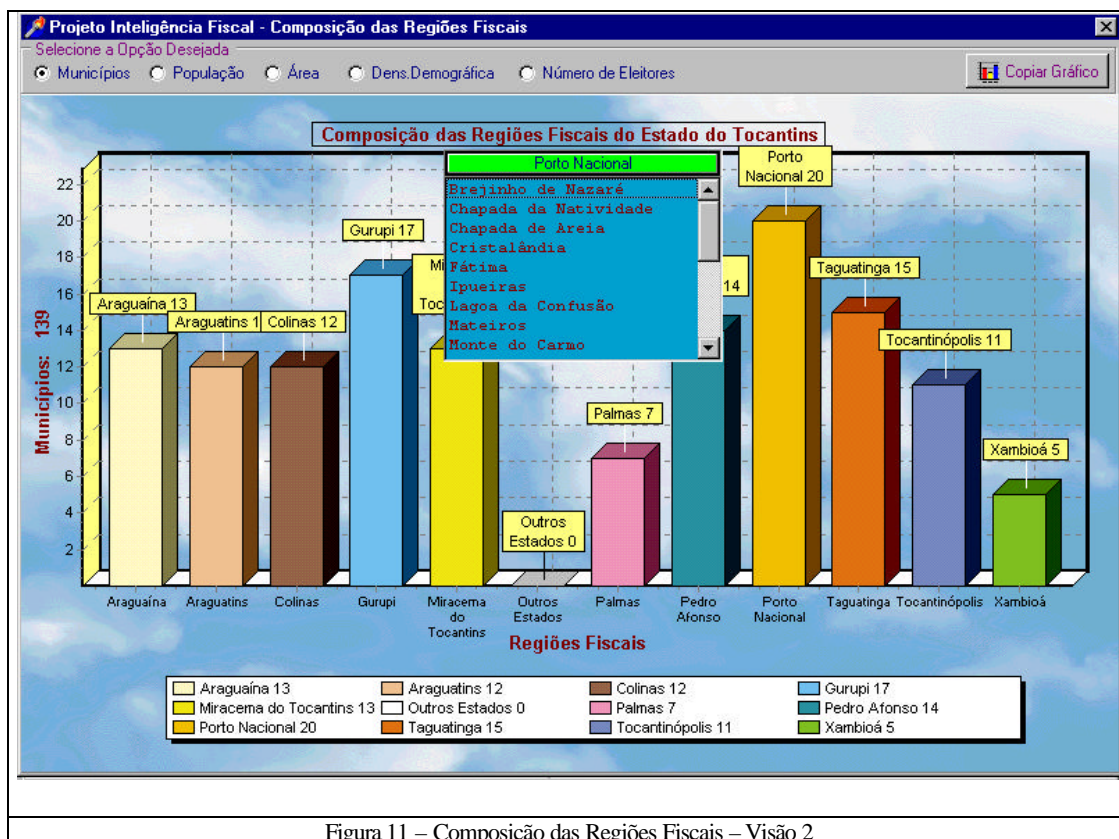


Figura 11 – Composição das Regiões Fiscais – Visão 2

As outras opções oferecidas por esta consulta se comportam de maneira análoga, tendo no entanto, um parâmetro a mais.

Isto significa que, se for selecionada, por exemplo, a População, o gráfico será refeito e a escala será automaticamente ajustada para este novo parâmetro.

Feito isto e pedindo-se um maior nível de detalhamento, aparecerão os valores das populações em ordem crescente, ao lado dos seus respectivos municípios, como mostra a figura 12.

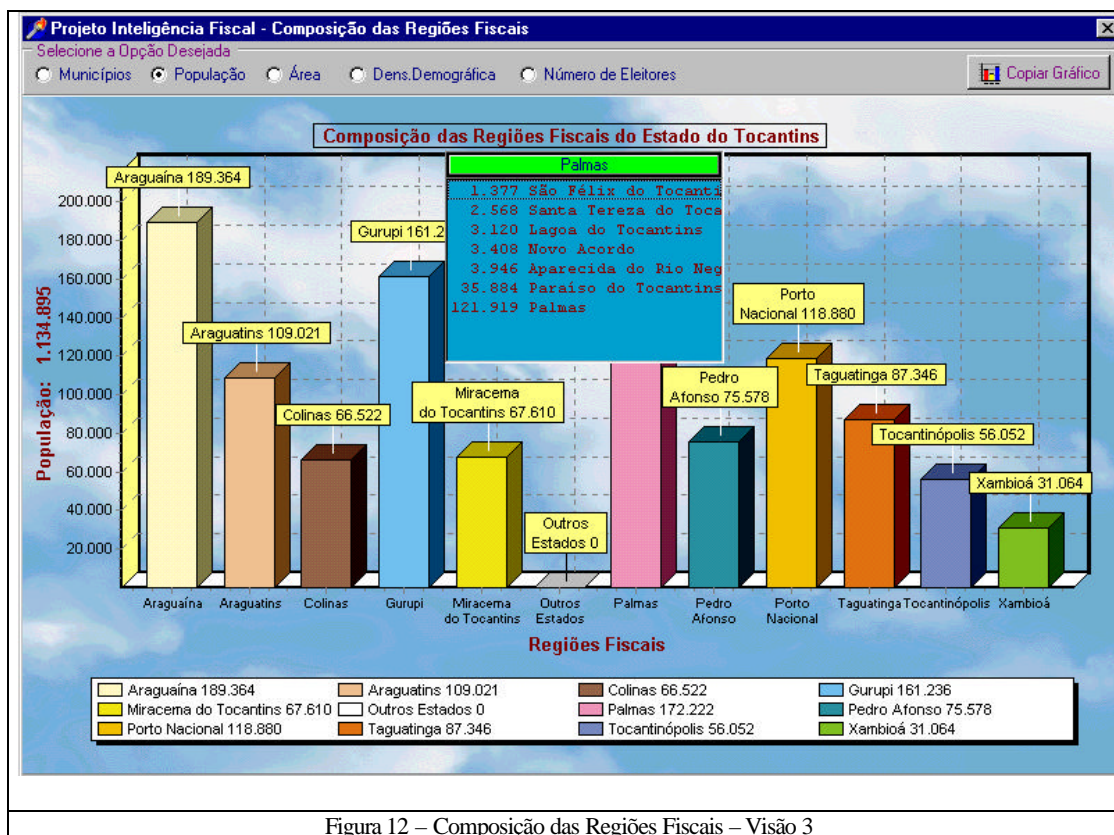


Figura 12 – Composição das Regiões Fiscais – Visão 3

Aqui se vê claramente, observando-se o nível global, que a população total do Estado é 1.134.895 habitantes. Num nível maior de detalhe, pode-se observar que a população da região fiscal de Araguaína é de 189.364 habitantes, a de Araguatins é de 109.021 habitantes, etc. Observando-se no nível máximo de detalhe, vê-se a população de todos os municípios que compõem a região fiscal de Palmas, selecionada para este exemplo. Ressalta-se aqui que estes são dados oficiais, obtidos diretamente do IBGE<sup>7</sup>

Ainda é possível saber a posição geográfica de qualquer município dentro do mapa do Estado, bastando para isto indicar o nome do município. Neste exemplo foi selecionado da figura 11, o município “Lagoa da Confusão”, resultando a figura 13.

<sup>7</sup> IBGE – Instituto Brasileiro de Geografia e Estatística

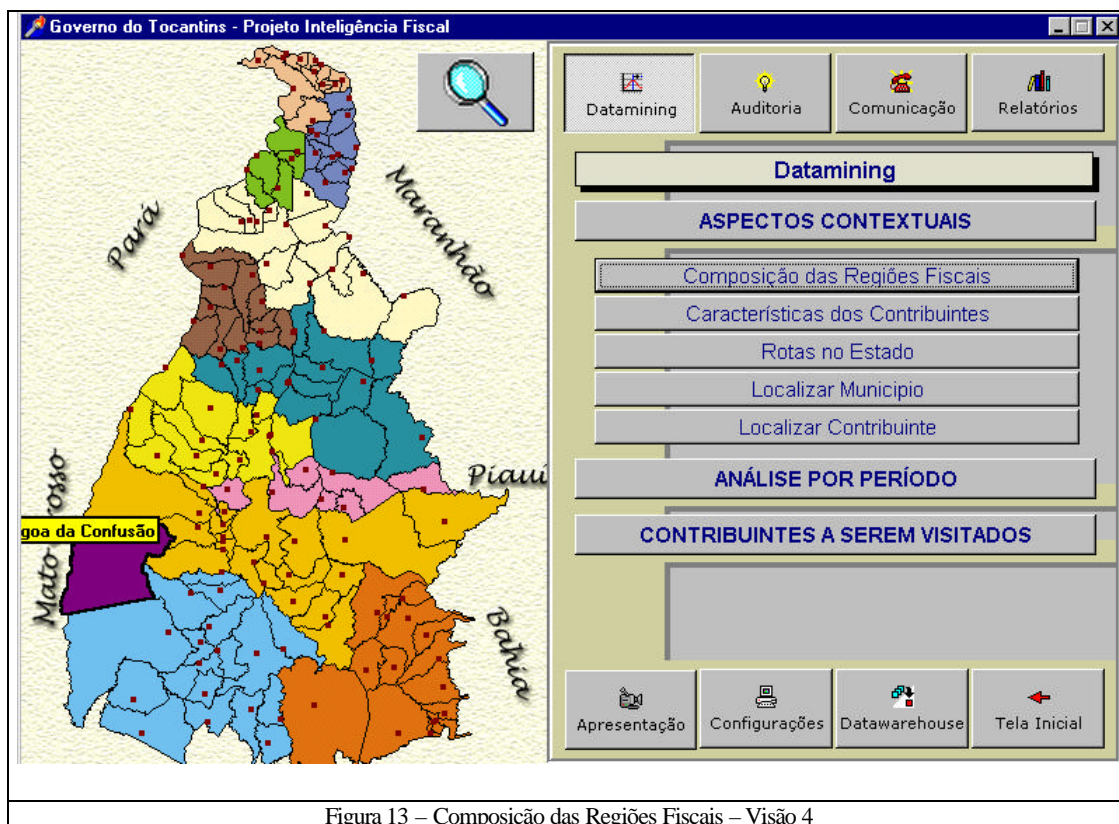


Figura 13 – Composição das Regiões Fiscais – Visão 4

O Sistema apresenta o resultado da consulta feita, destacando numa cor roxa, toda a área geográfica do município no mapa.



### 5.2.1.2 Características dos Contribuintes

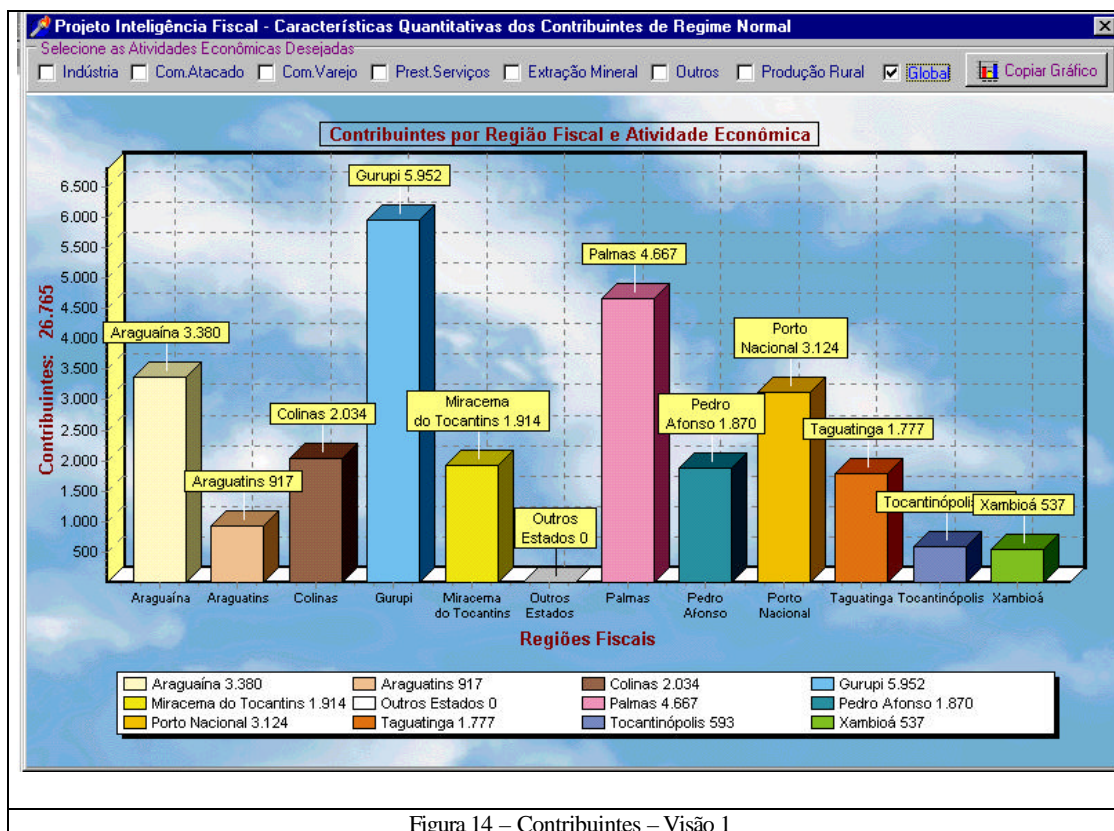


Figura 14 – Contribuintes – Visão 1

Observando-se a figura acima, pode-se ver que, de acordo com a opção selecionada (Global), foram apresentadas as quantidades de Contribuintes existentes em cada uma das regiões fiscais do Estado, independente do tipo de atividade econômica exercida por eles.

O eixo x (horizontal) apresenta as regiões fiscais do Estado, enquanto o eixo y (vertical) apresenta a quantidade de contribuintes que satisfaz os parâmetros selecionados. Ainda no eixo y é apresentada uma síntese do resultado da consulta.

Nesta consulta pode ser solicitado que os contribuintes sejam agrupados em função dos tipos de atividades econômicas exercidas por eles, e ainda, visando o conforto e facilidade de análise do usuário, pode ser solicitada a combinação destes tipos.

O número de barras apresentadas no gráfico para cada uma das regiões fiscais, depende da quantidade de atividades econômicas selecionadas, conforme os dois exemplos apresentados em seguida através das figuras 15, onde foram selecionadas simultaneamente as atividades econômicas indústria, comércio atacado, prestação de serviços e outros, e da figura 25, onde foram selecionadas produção rural e comércio a varejo.

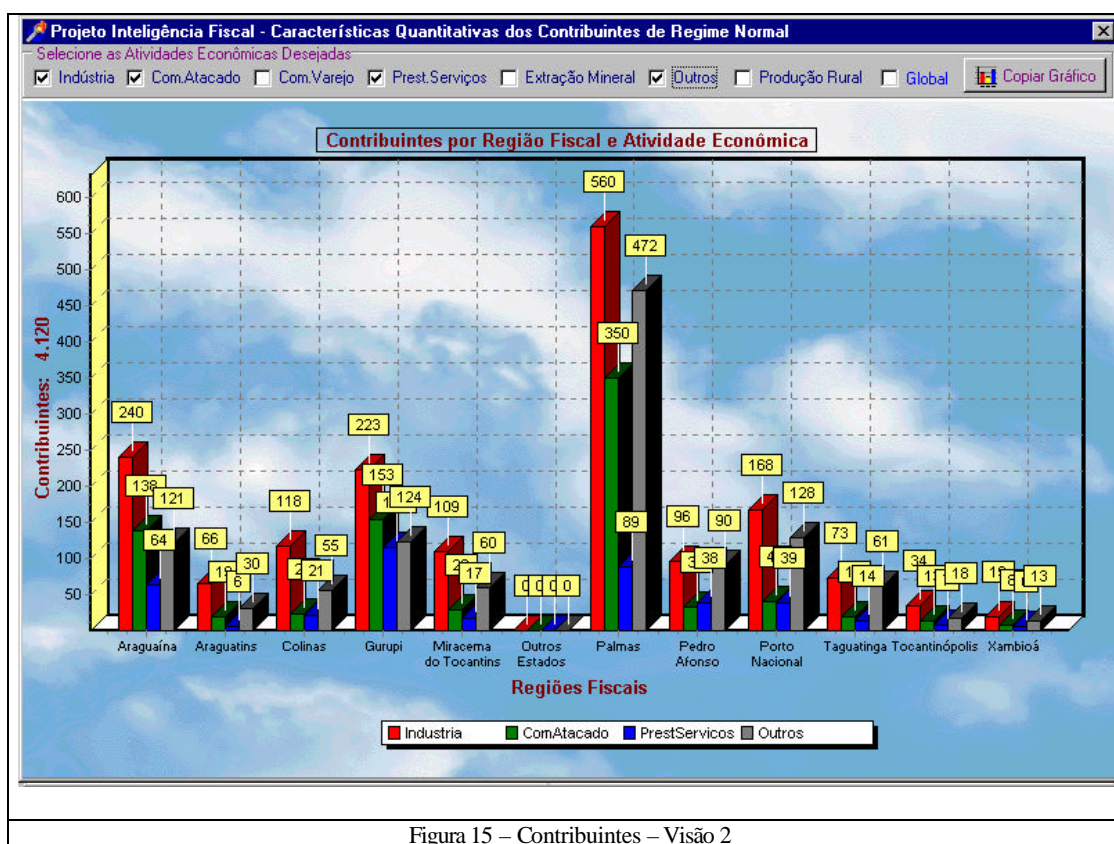


Figura 15 – Contribuintes – Visão 2

As quatro barras mostradas no gráfico para cada uma das regiões fiscais, refletem exatamente a solicitação feita através da consulta.

Cada barra representa uma atividade econômica, de acordo com a cor indicada na legenda, e apresenta a quantidade de contribuintes na região fiscal onde se encontra.

A escala do gráfico é dinâmica e se ajusta automaticamente em função do máximo entre todos os valores calculados, conforme pode ser observado nas figuras 14 a 16.

Na parte central da extrema esquerda do gráfico é apresentada ainda a quantidade total de contribuintes no Estado que trabalham com as atividades selecionadas.

As mesmas considerações se aplicam à figura 16, por similaridade.

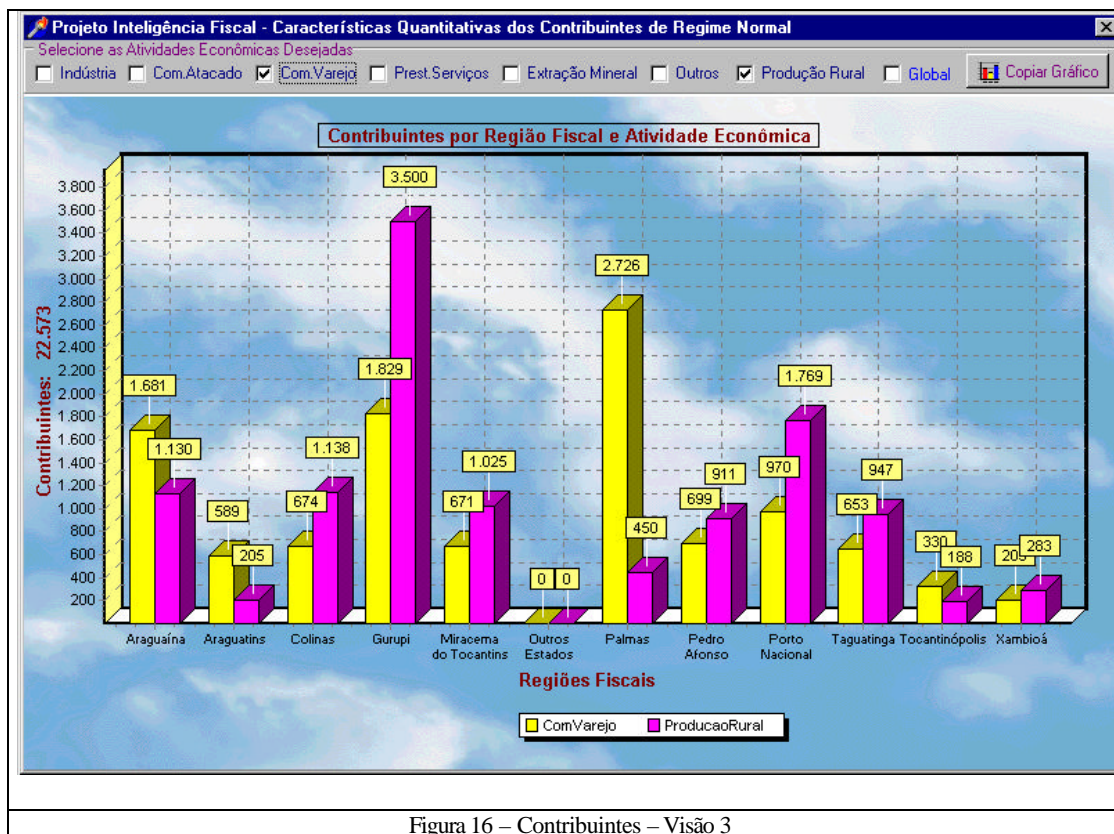


Figura 16 – Contribuintes – Visão 3

Ressalta-se a possibilidade de apresentação individual de qualquer uma das atividades econômicas, bem como a composição daquelas para as quais se deseja uma análise comparativa, o que implica em 128 combinações, nesta aplicação do modelo.

Além disso, a partir deste ponto, ainda é possível obter mais dois níveis de detalhamento (só existem mais dois níveis mesmo).

O primeiro deles para mostrar a quantidade de contribuintes por município dentro de uma determinada “região fiscal / atividade econômica” (Figura 17).



| Município                 | Número de Contribuintes |
|---------------------------|-------------------------|
| Aliança do Tocantins      | 157                     |
| Alvorada                  | 162                     |
| Araguaçu                  | 505                     |
| Cariri do Tocantins       | 178                     |
| Crixás do Tocantins       | 14                      |
| Dueré                     | 252                     |
| Figueirópolis             | 287                     |
| Formoso do Araguaia       | 482                     |
| Gurupi                    | 341                     |
| Jaú do Tocantins          | 128                     |
| Palmeirópolis             | 207                     |
| Peixe                     | 290                     |
| Sandolândia               | 137                     |
| São Salvador do Tocantins | 25                      |
| São Valério da Natividade | 89                      |

Figura 17 – Contribuintes – Visão 4

Nesta figura, vê-se os municípios da região fiscal selecionada (Gurupi), e neles a quantidade de contribuintes que atuam no ramo da atividade econômica indicada (Produção Rural).

O segundo nível de detalhamento se encarrega de apresentar a relação individual dos contribuintes do município selecionado, que atuam no ramo previamente indicado (vide figura 18).



Síntese de contribuintes em Crixás do Tocantins

Atividade: PRODUCAO RURAL

Pesquisar Contribuinte:

| Razão Social                    | Inscrição Estadual | CGC                | Atividade | En |
|---------------------------------|--------------------|--------------------|-----------|----|
| A M COSTA & CIA                 | 29 02 041148-3     | 37.583.457.0001-86 | 070108    | RC |
| ARMANDO NICOLA TARALLO          | 29 01 049229-7     | 00.027.874.9588-53 | 070203    | LC |
| CONDOMINIO CASA BRANCA          | 29 01 049250-5     | 00.003.545.5041-15 | 070108    | LC |
| CONDOMINIO FLOR DE GOIAS        | 29 01 026398-0     | 00.005.312.9068-91 | 070203    | RC |
| ERNANDO GLIENKE                 | 29 01 026382-4     | 00.005.344.6460-20 | 070107    | LT |
| ESTEVAO ALVES DA CUNHA          | 29 01 040256-5     | 00.009.974.5981-68 | 070108    | FA |
| GERALDO DAL BOM                 | 29 01 014666-6     | 00.013.645.8388-72 | 070203    | LC |
| HUMBERTO WALLAU                 | 29 01 040770-2     | 00.050.741.5860-53 | 070203    | LT |
| JOACY MADEIRA CRUZ              | 29 01 041166-1     | 00.000.802.6976-15 | 070203    | LC |
| JOSE SIMOES DE ALMEIDA SOBRINHO | 29 01 041179-3     | 00.084.895.8538-53 | 070203    | LT |
| MARCIO FRANCO VILAS BOAS        | 29 01 049225-4     | 00.081.499.4186-04 | 070108    | LC |
| MARCIO FRANCO VILAS BOAS        | 29 01 049232-7     | 00.081.499.4186-04 | 070108    | LC |
| MARCOLINO ARAUJO COSTA          | 29 01 011206-0     | 00.003.684.9036-04 | 070203    | RC |
| MARLENE FERNANDES COSTA         | 29 01 015788-9     | 00.009.433.3331-87 | 070203    | LC |

Figura 18 – Contribuintes – Visão 5

Na realidade, o modelo trabalha com quatro níveis de estratificação, onde o primeiro deles aborda o Estado como um todo, o segundo enfoca as regiões fiscais, o terceiro detalha por município e o último chega ao contribuinte. Neste último nível ainda existe a possibilidade de consultar rapidamente se um determinado contribuinte está ou não nesta relação.



### 5.2.1.3 Rotas no Estado

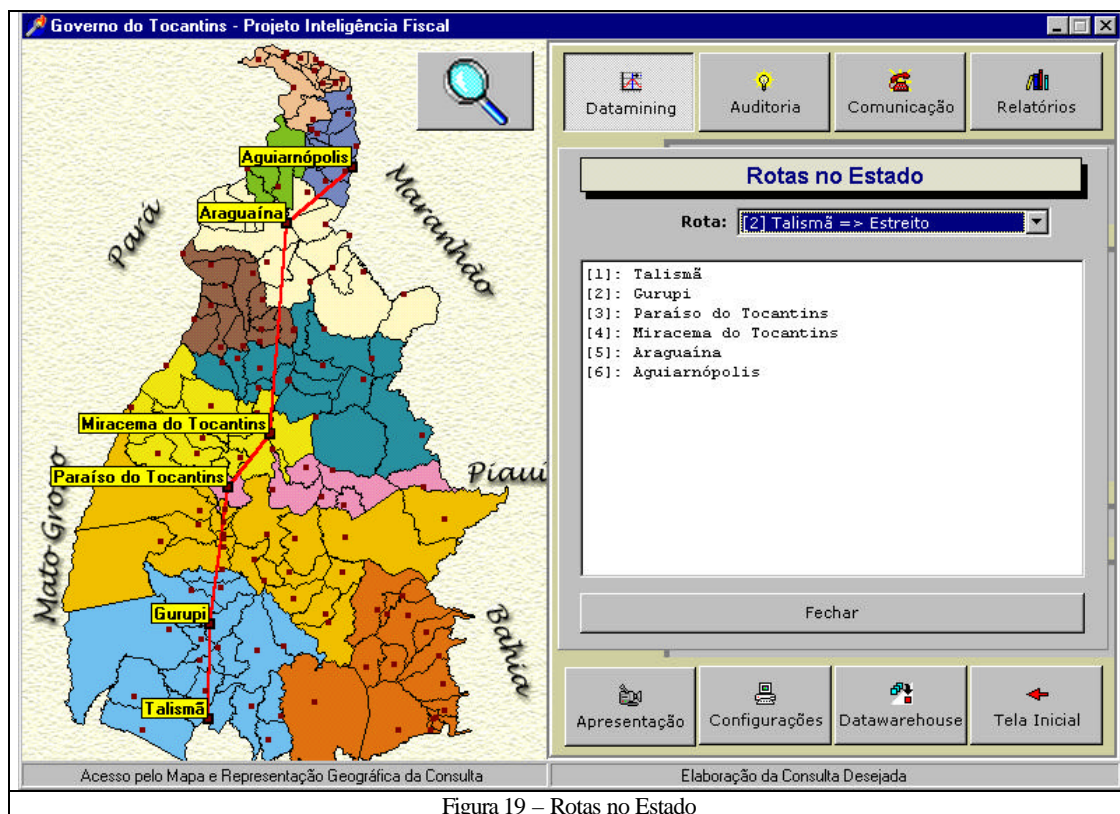


Figura 19 – Rotas no Estado

A figura 19 apresenta as diretrizes a serem seguidas pelo motorista, para ir de uma cidade a outra, utilizando as rodovias federais e estaduais que passam pelo Estado. Foi observado experimentalmente que, além de uma cortesia considerada muito útil pelos motoristas que trafegam pelo Estado, esta consulta ajudou a esclarecer dúvidas e confirmar dados fornecidos pelos próprios agentes fiscais.

### 5.2.1.4 Localizar Município

**Obter dados de um município**

Pesquisar Município:

| Nome                      | População | Nº de Eleitores |
|---------------------------|-----------|-----------------|
| Bandeirantes do Tocantins | 2.015     | 1.606           |
| Centenário                | 2.361     | 1.237           |
| Goianorte                 | 5.110     | 2.921           |
| Mateiros                  | 1.776     | 726             |
| Miranorte                 | 10.600    | 8.604           |
| Monte do Carmo            | 5.429     | 3.505           |
| Monte Santo do Tocantins  | 1.640     | 1.166           |
| Palmeirante               | 3.862     | 1.895           |
| Ponte Alta do Bom Jesus   | 4.445     | 2.395           |
| Ponte Alta do Tocantins   | 6.695     | 3.610           |

Nome:

Região Fiscal:  CEP:

Coletoria:

População:  Área:

Dens. Demográfica:  Nº de Eleitores:

Figura 20 – Localização de Municípios

Esta consulta têm o propósito de localizar rapidamente os dados básicos de qualquer um dos municípios, bem como a sua posição geográfica no mapa do Estado. Mostrou-se muito útil para os agentes fiscais, como também para aquele cidadão com interesse em conhecer o Estado por razões empresariais, turísticas, culturais, etc.

A pesquisa é feita informando incrementalmente letras que fazem parte do nome do município, como mostra a figura 20. A cada letra informada, são apresentados os municípios cujos nome a contém. Uma vez encontrado o resultado desejado, a sua posição geográfica no mapa é apresentada, nos moldes já vistos na figura 13.

### 5.2.1.5 Pesquisar Contribuinte

The screenshot shows a software window titled "Informações sobre os contribuintes". At the top, there is a search bar with the text "Contribuinte: carne". To the right of the search bar are three buttons: "Existe ?" (with a question mark icon), "Imprima" (with a printer icon), and "Fechar" (with a close icon). Below the search bar is a table with three columns: "Inscrição", "CGC", and "Razão Social". The table contains four rows of data, with the last row highlighted in blue. Below the table, there are several input fields for the selected contributor. The "CGC" field contains "26.936.484.0001-79" and the "Inscrição Estadual" field contains "29 02 033770-4". The "Razão Social" field contains "ABRAAO JORGE DA SILVA". The "Nome Fantasia" field contains "CASA DE CARNE CENTRAL". The "Endereço" field contains "AVE DAS PALMEIRAS 614". The "Bairro" field contains "CENTRO" and the "Município" field contains "Palmeirópolis". The "Telefone" field contains "00000000000" and the "CEP" field contains "77365-000". The "Porte" field contains "EMPRESA". The "Atividade" field contains "050104" and the "Situação" field contains "06" and "RECADASTRAMENTO". The "Início/Fim Ativ" field contains "07-05-1991". Below these fields, there is a section for the "Contador" with fields for "CGC" (26.936.484.0001-79), "Nome" (ENILVALDO LEAL DA SILVA), "Endereço" (AV. DAS PALMEIRAS N 1393), "Telefone" (00638861455), "Município" (Palmeirópolis), and "CRC" (608/TO).

| Inscrição      | CGC                | Razão Social             |
|----------------|--------------------|--------------------------|
| 29 03 056483-2 | 00.042.275.8731-04 | ABEL CLEITON DA SILVA    |
| 29 99 023250-3 | 00.898.163.0001-33 | ABERLARO PINHEIRO TORRES |
| 29 03 055639-2 | 00.022.385.2821-00 | ABRAAO BRAGA DE SOUZA    |
| 29 02 033770-4 | 26.936.484.0001-79 | ABRAAO JORGE DA SILVA    |

CGC: 26.936.484.0001-79 Inscrição Estadual: 29 02 033770-4

Razão Social: ABRAAO JORGE DA SILVA

Nome Fantasia: CASA DE CARNE CENTRAL

Endereço: AVE DAS PALMEIRAS 614

Bairro: CENTRO Município: Palmeirópolis

Telefone: 00000000000 CEP: 77365-000 Porte: EMPRESA

Atividade: 050104 CARNES E DERIVADOS, AVES PEIXES E DE OUTRO ANIMAIS,(CASA C

Situação: 06 RECADASTRAMENTO Início/Fim Ativ: 07-05-1991

Contador:

CGC: 26.936.484.0001-79 CRC: 608/TO

Nome: ENILVALDO LEAL DA SILVA

Endereço: AV. DAS PALMEIRAS N 1393

Telefone: 00638861455 Município: Palmeirópolis CEP: 77365-000

Figura 21 – Localização de Contribuintes

Esta consulta tem o propósito de localizar rapidamente os dados básicos de qualquer um dos contribuintes do Estado, independente da sua situação cadastral, isto é, mesmo que o contribuinte já tenha até encerrado a sua empresa, seus dados podem ser recuperados. A solicitação da pesquisa é feita fornecendo qualquer dado que se conheça sobre o contribuinte, mesmo que este dado esteja incompleto. No exemplo real da figura 21 foi fornecida a palavra carne, o que fez o modelo recuperar os contribuintes que trabalham com carne. Observou-se experimentalmente que esta consulta é de extrema utilidade para o setor específico de cadastro.

### **5.2.2 Análise dos Tributos por Período**

Este grupo está encarregado de analisar os dados relativos aos tributos estaduais propriamente ditos, para efeito de acompanhamento, verificação da qualidade, e previsão de receita. A SEF necessita que estas informações reflitam o resultado do mês de referência, já que a arrecadação é mensal; no entanto, a migração pode ocorrer em intervalos variáveis dentro deste período.

Enquadram-se neste grupo, os dados que possuem alta rotatividade.

O modelo forneceu todas as informações requisitadas pelo Usuário de acordo com o nível de acesso que lhe foi atribuído, possibilitando a visualização das mesmas em nível estadual, de região fiscal, de município e de contribuinte, além de ter possibilitado também a análise dos valores por atividade econômica e/ou combinação delas.

Descobriu-se, usando o modelo, que ainda existem muitos erros nestes dados, e que uma atenção especial deve ser dada para o tratamento dos mesmos.

A seguir são apresentadas consultas que pertencem a este grupo.

#### **5.2.2.1 Análise das Atividades Econômicas**

Informações geradas pelo modelo a partir dos dados financeiros das atividades econômicas.

O levantamento feito na fase inicial do projeto relacionou todas as informações que o usuário conseguiu lembrar e considerou necessárias ao seu trabalho. Visando sistematizar as solicitações, foi desenvolvida uma interface para esta parte do modelo, que consegue englobar todas estas consultas. (Figura 22)

## Análise de tributos por período - Interface genérica

The screenshot displays a web-based interface for tax analysis. It is organized into three main sections, each with a title bar and a list of checkboxes:

- Atividades Econômicas Desejadas:** Contains checkboxes for 'Indústria', 'Com. Varejo' (checked), 'Produção Rural', 'Com. Atacado' (checked), 'Extração Mineral', 'Global', 'Prest. Serviços', 'Outros', and 'Agrupar Ativ. Seleccionadas'.
- Informações Desejadas:** Contains checkboxes for 'Faturamento', 'ICMS Declarado' (checked), 'ICMS Arrecadado' (checked), 'Total Débitos', 'Total Créditos', and 'Arrecadação Per Capita'.
- Período:** Includes two date pickers: 'Início:' set to 'Outubro/1999' and 'Fim:' set to 'Dezembro/1999'. To the right of these is a checkbox for 'Totalizar no Estado' and a 'Pesquisar' button.

At the bottom of the interface, there is a caption: 'Figura 22 – Análise de tributos – Visão 1'.

Como pode ser visto na figura 22, a interface é composta assim:

Atividades Econômicas Desejadas: onde se assinalam as atividades que devem ser consideradas na análise.

Informações Desejadas: onde se assinalam os parâmetros que devem ser levados em consideração ao se analisar as atividades seleccionadas.

Período: onde se indicam o início e fim do intervalo de tempo dos dados que serão utilizados na análise.

Neste exemplo de consulta, foram seleccionadas as atividades económicas de Comércio Atacado e Comércio a Varejo, para serem analisadas conjuntamente em relação aos seus

respectivos ICMS delarado e ICMS Arrecadado, utilizando os dados do intervalo Outubro/1999 a Dezembro/1999.

O resultado está apresentado na Figura 23, a seguir:

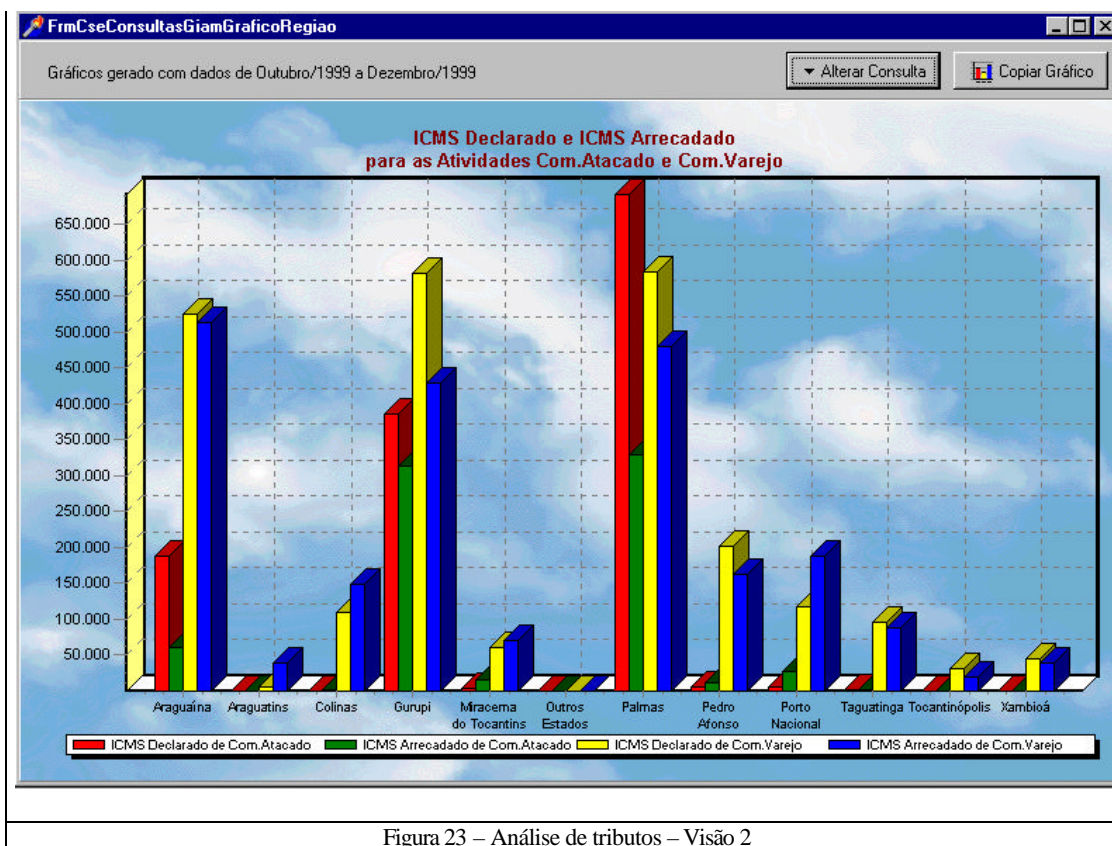


Figura 23 – Análise de tributos – Visão 2

Aqui, para cada região fiscal, foram desenhadas quatro barras, que, em sintonia com a consulta feita, representam, como indica a legenda:

Vermelho: icms declarado do comércio atacado.

Verde: icms arrecadado do comércio atacado.

Amarelo: icms declarado do comércio à varejo.

Azul: icms arrecadado do comércio à varejo.

Utilizando-se a interface apresentada anteriormente pode-se solicitar ainda:

Agrupamento das atividades selecionadas: para acumular os valores dos parâmetros especificados, para todas as atividades selecionadas, por região fiscal.

Totalização estadual: para obter o total estadual dos valores dos parâmetros das atividades selecionadas.

#### **5.2.2.2 Análise do movimento de Mercadorias**

Informações geradas pelo modelo que permitem o exame das mercadorias que chegam, saem ou circulam pelo Estado, sob todas as óticas necessárias a SEF.

A estratégia adotada aqui foi desenvolver uma interface de consulta que permite responder as todas as perguntas feitas pelos Usuários da SEF.

Esta interface está apresentada na Figura 24, juntamente com o resultado de uma consulta que mostra todas as notas fiscais que chegaram ao estado do Tocantins entre os dias 01 de outubro de 1999 e 30 de outubro de 1999.



**Análise do movimento de mercadorias**

**Origem das Mercadorias**

☐ Contribuinte (CGC)  
☒ Estado (UF)   
☐ Todos

**Destino das Mercadorias**

☐ Contribuinte (CGC)  
☒ Estado (UF)   
☐ Todos

**Período**

Início:   
 Fim:   
☐ Síntese

Fechar  
 Consultar  
 Imprimir  
 Ver Gráfico

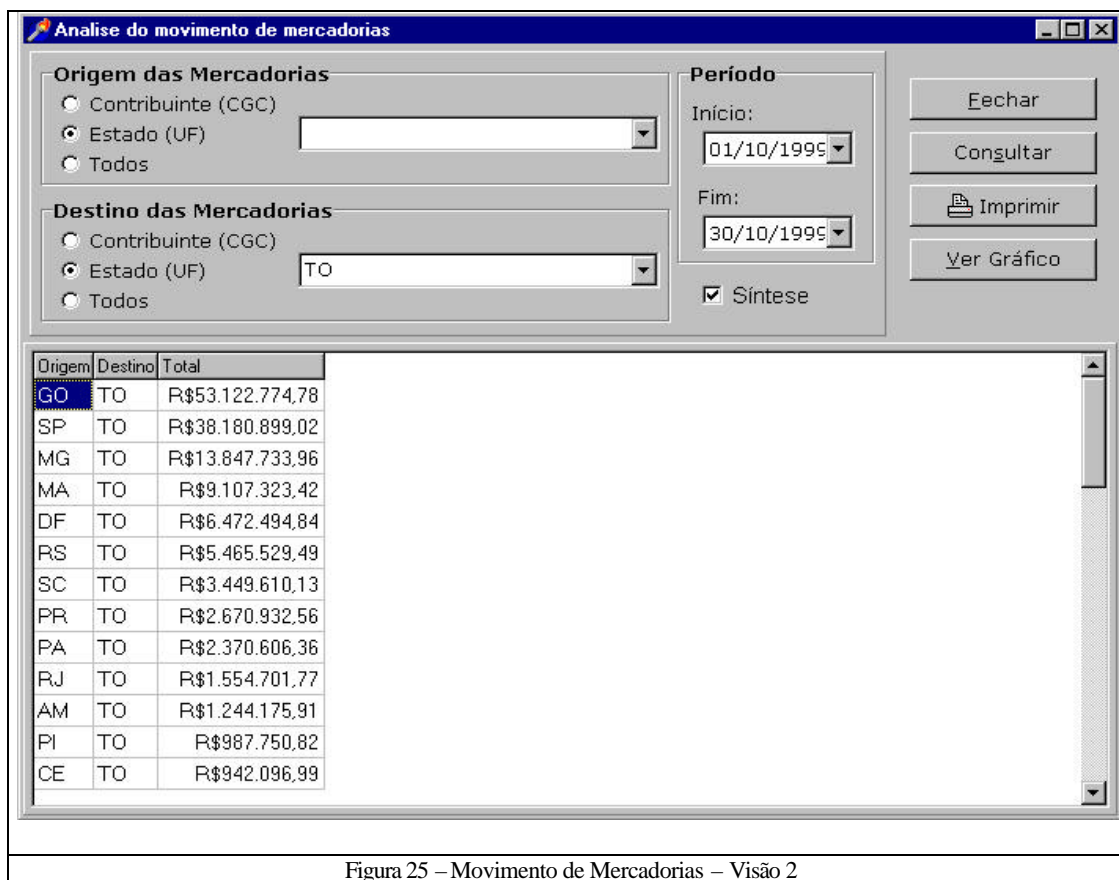
| Origem             | Valor        | Nota_Fiscal        | Data_Emissao | UF_Orig | RS_Origem                 |
|--------------------|--------------|--------------------|--------------|---------|---------------------------|
| 00.000.000.0018-30 | R\$66.774,19 | 00000000000549402  | 13-10-1999   | SP      | MINISTERIO DA INDUSTRIA C |
| 00.000.000.0018-30 | R\$66.338,71 | 00000000000554902  | 25-10-1999   | SP      | MINISTERIO DA INDUSTRIA C |
| 00.000.055.3185-33 | R\$13.500,00 | 00000000000057706  | 16-10-1999   | MA      | FRANCISCO MOURA MACED     |
| 00.000.102.0576-48 | R\$647,80    | 000000000002474302 | 18-10-1999   | GO      | PNEULANDIA - REFORMADO    |
| 00.000.213.3652-00 | R\$900,00    | 00000000017024406  | 15-10-1999   | PA      | MICHIKAZU TAKAKURA        |
| 00.000.374.5048-49 | R\$590,00    | 00000000000004902  | 05-10-1999   | SP      | AGUINALDO P DASILVA       |
| 00.000.404.5478-93 | R\$25.000,00 | 00000000078165702  | 27-10-1999   | GO      | CARLOS ROBERTO DE OLIVE   |
| 00.000.462.9421-04 | R\$7.380,00  | 00000000163264402  | 08-10-1999   | PA      | LINDOLFO BENTO PEREIRA.   |
| 00.000.462.9421-04 | R\$7.380,00  | 00000000163268802  | 08-10-1999   | PA      | LINDOLFO BENTO PEREIRA.   |
| 00.000.462.9421-04 | R\$7.380,00  | 00000000163266602  | 08-10-1999   | PA      | LINDOLFO BENTO PEREIRA.   |
| 00.000.462.9421-04 | R\$7.380,00  | 00000000163267702  | 08-10-1999   | PA      | LINDOLFO BENTO PEREIRA.   |
| 00.000.462.9421-04 | R\$7.000,00  | 00000000163272202  | 08-10-1999   | PA      | LINDOLFO BENTO PEREIRA.   |
| 00.000.462.9421-04 | R\$7.000,00  | 00000000163271102  | 08-10-1999   | PA      | LINDOLFO BENTO PEREIRA.   |

Figura 24 – Movimento de Mercadorias – Visão 1

Como pode ser visto na figura acima, o modelo de consulta projetado permite especificar origem e destino das mercadorias e todas as opções possíveis para ambos. Desta forma foi possível responder a todas as perguntas e sanar todas as dúvidas apresentadas pelos Usuários da SEF, que se disseram plenamente satisfeitos com os resultados.

Nesta mesma interface, ainda é possível especificar que seja apresentada uma síntese da consulta feita, o que, no caso deste exemplo, apresentou o resultado mostrado na Figura 25.





Esta síntese apresenta, em ordem decrescente de relevância tributária, os valores totais de mercadorias provenientes de cada um dos Estados do Brasil para um determinado Estado escolhido pelo Usuário, num determinado intervalo de tempo. Segundo os profissionais da SEF: “Este modelo nos fornece todas as informações que precisamos para o planejamento das nossas ações de fiscalização”.

### 5.2.3 Contribuintes a Serem Visitados

Este grupo está encarregado de fazer as análises que conduzem à indícios de evasão fiscal, baseados nos quais se planejam as visitas produtivas dos fiscais às empresas; eles recebem a relação das empresas localizadas nas suas respectivas áreas de atuação, usualmete um ou mais municípios da região fiscal onde trabalham e, para cada uma delas, os indícios que devem ser verificados *in loco*.

As consultas que fazem parte deste grupo, são:

- Omissos de Giam: relação dos contribuintes que deveriam e não entregaram a Giam; esta relação aparece grupada por região fiscal/município, e contém os seguintes atributos: inscrição estadual, cgc/cpf, mês e ano da ocorrência, razão social, nome fantasia, endereço, código do município, código do grupo de atividades, código da atividade, código do status, código do porte e código da região fiscal do contribuinte.
- Imposto devido declarado e não pago: relação dos contribuintes que declararam mas não pagaram ou pagaram a menos o imposto devido, informado na Giam; esta relação aparece grupada por região fiscal/município, e contém os seguintes atributos: inscrição estadual, cgc/cpf, mês e ano da ocorrência, razão social, nome fantasia, endereço, código do município, código do grupo de atividades, código da atividade, código do status, código do porte e código da região fiscal do contribuinte.
- Indícios de evasão fiscal: relação dos contribuintes que apresentam indícios de evasão fiscal; esta relação aparece grupada por região fiscal/município, contendo os seguintes atributos: inscrição estadual, cgc/cpf, razão social, nome fantasia, código e descrição do indício encontrado, mês e ano da ocorrência, endereço, código do

município, código do grupo de atividades, código da atividade, código do status, código do porte e código da região fiscal do contribuinte.

Foi desenvolvida uma interface única para solicitação das consultas e apresentação dos resultados deste grupo, devido à característica de similaridade existente entre os parâmetros utilizados por elas (Figura 26).

The screenshot shows a software window titled "Indícios de sonegação...". At the top, it displays "Contribuintes Suspeitos (que não se encaixam nos perfis de suas atividades)" and "Nº de contribuintes localizados: 0". Below this, there are several input fields and a list:

- Município:** A dropdown menu with "Palmas" selected.
- Região Fiscal:** A dropdown menu with "Palmas" selected. A list of options is visible: "Miracema do Tocantin", "Outros Estados", "Palmas", "Pedro Afonso", "Porto Nacional", "Taguatinga", "Tocantinópolis", and "Xambioá".
- Mês:** A dropdown menu with "Outubro/1999" selected.
- Contribuinte (Inscrição):** An empty text input field.

At the bottom right, there are two buttons: "Consultar" (with a magnifying glass icon) and "Imprimir" (with a printer icon). The main area of the window is a large empty box, likely for displaying the search results.

Figura 26 – Contribuintes a serem visitados

Neste exemplo está sendo solicitada a relação dos contribuintes da região fiscal de Palmas, sediados no município de Palmas e que apresentaram indícios de evasão fiscal no mês de outubro de 1999. A confidencialidade destas informações impede que os contribuintes sejam divulgados.

A interface permite a consulta nos quatro níveis possíveis: estadual, região fiscal, município e contribuinte.

Observou-se que o modelo, utilizando um determinado conjunto de dados escolhido por profissionais da SEF, descobriu uma relação muito grande de indícios, que, após analisados pelos conhecedores das empresas e do contexto, mostraram-se verdadeiros à luz dos dados que foram utilizados. Este fato validou a estratégia do modelo, dando-lhe credibilidade e motivou os profissionais da Fazenda envolvidos. Como ainda existem erros nos dados que estão sendo utilizados, os esforços estão todos concentrados no acerto dos mesmos.

Os principais indícios detectados pelo modelo foram:

- Omissos de GIAM
- Valor pago inferior ao valor devido
- Fora do perfil valor declarado/faturamento para a atividade econômica
- Fora do perfil valor pago / faturamento para a atividade econômica
- Fora do perfil débito/crédito para a atividade econômica
- Fora do perfil consumo energia elétrica/faturamento para a atividade econômica
- Fora do perfil comissões/faturamento para a atividade econômica
- Fora do perfil total das despesas/faturamento para a atividade econômica.
- Novos indícios poderão surgir e ser incorporados à base de indícios do modelo.

### **5.3 Conclusão**

Este modelo pode ser aplicado em qualquer uma das SEF's dos estados brasileiros, fazendo-se as adaptações necessárias ao atendimento das suas peculiaridades.

As 161 entidades relacionadas no modelo, mostraram-se abrangentes o suficiente para contemplar as necessidades de informações da SEF onde foi aplicado, e os algoritmos descobriram indícios que foram posteriormente verificados e comprovados “in loco”.

O profissional da Fazenda tem agora uma poderosa ferramenta de trabalho e não precisa depender de terceiros para fazer consultas, elaborar sua própria análise e obter os relatórios de interesse.

## 6 CONCLUSÃO E RECOMENDAÇÕES

A coibição da evasão dos tributos devidos mostrou-se como a alternativa mais adequada para aumentar a arrecadação da receita dos estados brasileiros, sem instituir novos impostos ou majorar a carga tributária dos contribuintes.

Como a administração fazendária tem uma série de outras atribuições e necessidades, apresentadas no decorrer deste trabalho, diversas ações foram postas em prática visando dotá-la de condições e instrumentos capazes de propiciar uma política fiscal transparente, priorizando a receita tributária e o aprimoramento da gestão das finanças públicas.

Estas ações, parte integrante do presente modelo, foram submetidas a rigorosas averiguações práticas e, com as adaptações decorrentes delas, mostraram-se plenamente satisfatórias, ficando assim comprovada a eficácia e viabilidade das mesmas.

Relembrem-se abaixo as questões tratadas pelo presente modelo, e as correspondentes ações, por área:

- **Organização e Gestão:** Desenvolvimento do planejamento estratégico das necessidades de informação, materializado através de um modelo de dados (*datawarehouse*), junto com o seu metadado, capaz de contemplar todos os atributos de interesse para a tomada de decisões.
- **Tecnologia da Informação:** Utilização da tecnologia Cliente-Servidor, Especificação dos equipamentos apropriados para o bom desempenho do modelo, Transferência de tecnologia através da participação do pessoal diretamente envolvido. Estes tópicos, juntamente com outros que fogem do escopo deste trabalho, devem fazer parte do Plano Diretor de Informática.

- **Cadastro dos Contribuintes:** Obtenção de dados dos contribuintes provenientes da Junta Comercial e da Companhia de Energia Elétrica, visando melhorar o referido cadastro. Alguns outros problemas, como por exemplo, o da falta de integridade referencial, foram identificados automaticamente pelo modelo. Foi também alvo de análise na busca de melhoria do cadastro, o sistema SINTEGRA<sup>8</sup>, que infelizmente não ajudou muito, pois continha os mesmos dados básicos já existentes nas SEF's.
- **Mercadorias:** Identificação e apresentação em níveis gerenciais adequados, dos problemas detectados nas notas fiscais originárias, destinatárias ou que circulam pelo Estado, facilitando assim o trabalho de verificação e providências decorrentes dele.
- **Fiscalização:** Identificação automática dos contribuintes com indícios de evasão fiscal, apresentando-os de forma a facilitar o planejamento da ação dos fiscais, que recebem em suas respectivas áreas de atuação, a relação dos contribuintes que devem ser visitados, juntamente com os indícios previamente descobertos pelo modelo.
- **Arrecadação:** Apresentação de dados e informações de uma maneira bastante flexível, de forma a permitir o acompanhamento e a análise do comportamento da receita global, ou por grupo de atividade econômica. Na medida em que a quantidade de documentos recebidos eletronicamente for aumentando, irá se tornando mais viável o desenvolvimento de um programa de previsão de arrecadação cada vez mais perto da realidade.

---

<sup>8</sup> SINTEGRA/ICMS - Sistema Integrado de Informações sobre Operações Interestaduais com Mercadorias e Serviços, que consiste num encadeamento de procedimentos administrativos e sistemas computacionais adotado em diversos Estados Brasileiros, visando simplificar e uniformizar as obrigações dos contribuintes relacionadas com as informações econômico-fiscais.

- **Contencioso Administrativo:** Gerenciamento dos dados sobre os processos e sobre o cadastro amigável da dívida ativa. Com base na análise destes dados, que refletem a situação atual, poderão ser reestruturados os procedimentos relativos ao contencioso administrativo tributário, visando uma maneira para melhorar a montagem dos processos (peças necessárias), o julgamento do contencioso (instâncias), os prazos a serem observados no andamento do processo, as medidas legais e regulamentares a serem implementadas, etc.

Foi de grande valia a divulgação e discussão aberta deste modelo nas delegacias, nas coletorias, nos postos fiscais, nas administrações centralizadas e descentralizadas, como também no Conselho Estadual dos Contribuintes.

Em todos estes lugares notou-se que, na medida em que as soluções iam surgindo, a ansiedade dos profissionais que participaram das discussões e da busca de soluções ia crescendo, fato que reflete bem o estado de espírito da grande maioria do povo brasileiro diante do quadro atual de descontrole e impunidade, mostrado a toda hora pelos veículos de comunicação.

Notou-se também, que ao se demonstrar controle sobre a situação, está se fazendo uma política preventiva para a evasão fiscal e que este é o caminho mais curto para o aumento da arrecadação tributária.

O Estado de Santa Catarina, que participa com 3,4% da arrecadação tributária do país, tem uma evasão mensal aproximada de R\$ 78.000.000,00 (setenta e oito milhões de reais).

Este trabalho contribui de maneira decisiva para o aprimoramento da gestão das finanças públicas, através do efetivo controle dos processos de arrecadação de tributos,



fornecendo à administração tributária os elementos que lhe permitem acompanhar de perto todas as ações dos contribuintes, detectando e informando as distorções encontradas.

Isto foi possível graças ao estudo e compreensão dos problemas reais enfrentados pela administração tributária, onde foi fundamental a participação dos profissionais que conhecem profundamente este campo, e a aplicação conveniente de técnicas modernas apresentadas no decorrer do trabalho, adaptadas para a solução de tais problemas.

Estas conclusões não poderiam ser finalizadas sem dizer que a Universidade está estreitando cada vez mais os seus laços com a Comunidade, e que foi muito gratificante tê-la ajudado nesta nobre missão.

Recomendações para futuros trabalhos:

A inclusão de estudos econômicos tributários, na formulação de uma política fiscal que defina e utilize indicadores do comportamento dos setores econômicos, deve ser considerado como um tópico complementar, de muita utilidade para as SEF's.

Aplicação da Inteligência Fiscal nas Prefeituras, visando otimizar a administração de recursos e aumentar da arrecadação dos tributos municipais, apoiando a Lei de Responsabilidade Fiscal – LRF (lei 101 de 04/05/2000).

A incorporação do IPVA<sup>9</sup> na determinação da receita.

Maior investimento no estudo da interação com o SINTEGRA.

Pesquisa de novas alternativas de melhorias para a gestão das finanças públicas.

Desenvolvimento de uma solução Data Webhouse (KIMBALL, Jul 13, 1999), (KIMBALL, Nov 16, 1999), (KIMBALL, Jun 5, 2000)

---

<sup>9</sup> IPVA – Imposto sobre propriedade de veículos automotores

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, R., GHOSH, S., IMIELINSKY, T., IYER, B. and SWANI, A., **An Interval Classifier for Database Mining Application**, Proc. 18<sup>th</sup> Int'l Conf. Very Large Data Bases, pp. 560-573, Aug. 1992.

AGRAWAL, R., IMIELINSKI, T. and SWAMI A., **Database Mining: A Performance Perspective**, IEEE Transactions on Knowledge and Data Engineering., pp. 914-925, Dec. 1993.

AGRAWAL, R., MEHTA, M., SHAFER, J., SRIKANT, R., ARNING, A. and BOLLINGER, T., **The QUEST Data Mining System**, Proc. Int'l Conf. Data Mining and Knowledge Discovery (KDD '96), pp. 244-249, Portland, Ore., Aug. 1996.

AGRAWAL, R. and SRIKANT, R., **Fast Algorithms for Mining Association Rules in Large Databases**, Proc. 20<sup>th</sup> Int'l Conf. Very Large Data Bases, pp. 478-499, Sept. 1994.

ARMSTRONG, R., **Data Warehousing: Clearing the Confusion**, <http://www.ncr.com>, Mar. 2000.

BERRY, M.J.A. and LINOFF, G., **Data Mining Techniques**, John Wiley & Sons, New York, 1997.

BERSON, A., **Data Warehousing, Data Mining & OLAP**, McGraw-Hill, USA, 1997

BORT, J., **The Wiser, Gentler Data Warehouse**, <http://www.sunworld.com/unixinsideronline>, Jan. 2001.

BRACHMAN, R. and ANAND, T., **The process of Knowledge Discovery in Databases: A Human-Centered Approach**, Advances in Knowledge Discovery and Data Mining, pp. 37-58, AAAI/MIT Press, 1996.

BRACKETT, Michael H., **The Data Warehouse Challenge**, John Wiley & Sons, 1996.

BRAZ, Eugênio R.C., **Solução heurística para o problema do caixeiro viajante**. Departamento de Engenharia de Produção – UFSC, Florianópolis/Santa Catarina, 1980.

BREIMAN, L., FRIEDMAN, J. OLSHEN, R. and STONE, C., **Classification of Regression Trees**, Wadsworth, 1984.

CHEN, Ming-Syan, HAN, Jiawei and YU, Philip S., **Data Mining: An Overview from a Database Perspective**, IEEE Transactions on Knowledge and Data Engineering, Vol.8, N°6, December 1996.

FAYYAD, U.M., PIATETSKY-SHAPIO, G., SMITH, P., and UTHURUSAMY, R., **Advances in Knowledge Discovery and Data Mining**, AAAI/MIT Press, 1996.

FIRESTONE, J.M., Architectural Evolution in Data Warehousing and Distributed Knowledge Management Architecture, <http://www/dkms.com>, Oct. 2000.

GAINS B.R., **Transforming Rules and Trees into Comprehensive Knowledge Structures**, Advances in Knowledge Discovery and Data Mining, pp. 205-228, AAAI/MIT Press, 1996.

GIOVINAZZO, W.A., **Object-Oriented Data Warehouse Design – Building a Star Schema**, Prentice Hall, New Jersey, 2000.

GUPTA, A., HARINARAYAN, V. and QUASS, D., **Aggregate –Query Processing in Data Warehousing Environment**, Proc. 21<sup>th</sup> Int'l Conf. Very Large Data Bases, pp. 358-369, Zurich, Sept., 1995.

GAMMIL, P., Reading, Writing and Warehousing, DBMS, Dec. 1996.

GORKI S., CARVALHO A., **Tecnologias de Redes**. Book Express Ltda, Rio de Janeiro, 1998.

- HAIR J., JOSEPH F., **Multivariate data analysis**, Prentice-Hall, Upper Saddle River, 5 ed., New Jersey, 1998
- HAN, J., CAI, Y. and CERCONE, N., **Data-Driven Discovery of Quantitative Rules in Relational Databases**, IEEE Transactions on Knowledge and Data Engineering, vol.5, pp. 29-40, 1993.
- HAN, J. and FU, Y., **Discovery of Multiple-Level Association Rules from Large Databases**, Proc. 21<sup>th</sup> Int'l Conf. Very Large Data Bases, pp. 420-431, Sept. 1995.
- HAN, J. and FU, Y., **Exploration of the Power of Attribute-Oriented Induction in Data Mining**, Advances in Knowledge Discovery and Data Mining, pp. 399-421, AAAI/MIT Press, 1996.
- HAN, J., FU, Y., WANG, W., CHIANG, J., GONG, W., KOPERSKI, K., LI, D., LU, Y., RAJAN, A., STEFANOVIC, N., XIA, B. and ZAIANE, O.R., **A System for Mining Knowledge in Large Relational Databases**, Proc. Int'l Conf. Data Mining and Knowledge Discovery (KDD '96), pp. 250-255, Portland, Ore., Aug. 1996.
- HARINARAYAN, V., ULLMAN, J.D. and RAJARAMAN, A., **Implementing Data Cubes Efficiently**, Proc. ACM SIGMOD Int'l Conf. Management Data, pp. 205-216, Montreal, Canada, June 1996.
- HARRISON, T. H., **Intranet data warehouse**, São Paulo, Berkeley Brasil, 1998.
- INMON, W.H., **Como Construir o Data Warehouse**. Editora Campus Ltda, 1997.
- JOHNSON, R.A. and WICHERN, D. W., **Applied multivariate statistical analysis**, Prentice-Hall, 4. ed., New Jersey, 1998.
- KELLY, T.J., **Dimensional Data Modeling**, <http://www.sybase.com>, Nov. 2000.
- KIGHT, B., The Smart Way to Build a Data Warehouse, Datamation, Oct. 1996.

KIMBALL, R., **Is Your Dimensional Data Warehouse Expressive?** (The final eight criteria for comparing your systems), Intelligent Enterprise Magazine, May 15, 2000.

KIMBALL, R., **The Data Webhouse Has No Center** (Facing the challenges of profoundly distributed data warehouse design), Intelligent Enterprise Magazine, July 13, 1999.

KIMBALL, R., **Welcoming the Packaged App** (Make sure your packaged application is part of your data Webhouse), Intelligent Enterprise Magazine, June 5, 2000.

KIMBALL, R., **Working in Web Time** (An Architecture for the Webhouse), Intelligent Enterprise Magazine, November 16, 1999.

KIMBALL, R., REEVES, L., ROSS, M and THORNTHWAITE, W., **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses.** John Wiley & Sons, New York, 1998.

KIMBALL, R., **The Data Warehouse Toolkit. Practical Techniques for Building Dimensional Data Warehouses,** John Wiley & Sons, New York, 1996.

KLOSGEN, W., **Explora: A Multipattern and Multistrategy Discovery Assistant,** Advances in Knowledge Discovery and Data Mining, pp. 249-271, AAAI/MIT Press, 1996.

MARQUES, P.E., **Direito Tributário,** Apostilas Solução, São Paulo, Brasil, 1997.

MATHEUS, C.J., PIATETSKY-SHAPIO, G. and McNEIL D., **Selecting and Reporting What is Interesting: The KEFIR Application to Health Care Data,** Advances in Knowledge Discovery and Data Mining, pp. 495-516, AAAI/MIT Press, 1996.

MEHTA, M., AGRAWAL, R. and RISSANEN, J., **SLIQ: A Fast Scalable Classifier for Data Mining**, Proc. Int'l Conf. Extending Database Technology (EDBT'96), Avignon, France, Mar. 1996.

OLIVEIRA, E.A., **Artigo sobre Tributação**, Auditor Tributário da Secretaria da Fazenda, Pernambuco, Brasil, 1998.

PARK, J.S., CHEN, M.S. and YU, P.S., **An Effective Hash Based Algorithm for Mining Association Rules**, Proc. ACM SIGMOD, pp.175-186, May. 1995.

PIATETSKY-SHAPIO, G., **Discovery, Analysis, and Presentation of Strong Rules**, Knowledge Discovery in Databases, pp. 229-238, AAAI/MIT Press, 1991.

PNAFE Programa Nacional de Apoio à Administração Fiscal para os Estados Brasileiros, Governo do Estado do Tocantins, Secretaria da Fazenda

PPGEP **Inteligência Fiscal** – Sistema Inteligente para Auditoria Fiscal, 1998

QUINLAN, J.R., **Induction of Decision Trees**, Machine Learning, vol.1, pp.81-106, 1986.

QUINLAN, J.R., **C4.5: Programs for Machine Learning**, Morgan Kaufmann, 1993.

RICMS Regulamento do imposto sobre operações relativas à circulação de mercadorias e sobre prestações de serviços de transporte interestadual e intermunicipal e de comunicação (icms), Decreto nº 13.640, de 13 de novembro de 1997.

SELFIDGE, P.G., SRIVASTAVA and WILSON, L.O., **IDEA: Interactive Data Exploration and Analysis**, Proc. ACM SIGMOD Int'l Conf. Management Data, pp. 24-34, Montreal, Canada, June. 1996.

SIMON, Alan R., Strategic Database Technology: Management for the year 2000. Morgan Kaufmann Publishers, Inc, 1995.

SRIKANT, R. and AGRAWAL, R., **Mining Generalized Association Rules**, Proc. 21<sup>th</sup> Int'l Conf. Very Large Data Bases, pp. 407-419, Sept.. 1995.

SRIKANT, R. and AGRAWAL, R., **Mining Quantitative Association Rules in Large Relational Tables**, Proc. ACM SIGMOD Int'l Conf. Management Data, pp. 1-12, Montreal, Canada, June.. 1996.

WEISS, S.M. and KULIKOWSKI, C.A., Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems, Morgan Kaufmann, 1991.

WIDOM, J., **Research Problems in Data Warehousing**, Proc. Fourth Int'l Conf. Information and Knowledge Management, pp. 25-30, Baltimore, Nov. 1995.

YAN, W.P. and LARSON, P., **Eager Aggregation and Lazy Aggregation**, Proc. 21<sup>th</sup> Int'l Conf. Very Large Data Bases, pp. 345-357, Zurich, Sept., 1995.

ZIMMER, H., **Data Warehousing: Are You on a Path to Success or Failure?**, <http://www.dw-institute.com/whatworks10/lessons>, Jan. 2001.

## ANEXO I

### CAPÍTULO I

#### DO SISTEMA TRIBUTÁRIO NACIONAL

##### Seção I

##### Dos Princípios Gerais

Art. 145. A União, os Estados, o Distrito Federal e os Municípios poderão instituir os seguintes tributos:

I – impostos;

II – taxas, em razão do exercício do poder de polícia ou pela utilização, efetiva ou potencial, de serviços públicos específicos e divisíveis, prestados ao contribuinte ou postos à sua disposição;

III – contribuição de melhoria, decorrente de obras públicas.

&1º. Sempre que possível, os impostos terão caráter pessoal e serão graduados segundo a capacidade econômica do contribuinte, facultado à administração tributária, especialmente para conferir efetividade a esses objetivos, identificar, respeitados os direitos individuais e nos termos da lei, o patrimônio, os rendimentos e as atividades econômicas do contribuinte.

&2º. As taxas não poderão ter base de cálculo própria de impostos.

Art. 146. Cabe a lei complementar:

I – dispor sobre conflitos de competência, em matéria tributária, entre a União, os Estados, o Distrito Federal e os Municípios;

II – regular as limitações constitucionais ao poder de tributar;

III – estabelecer normas gerais em matéria de legislação especialmente sobre:



Definição de tributos e de suas espécies, bem como, em relação aos impostos discriminados desta Constituição, a dos respectivos fatos geradores, bases de cálculo e contribuintes;

Obrigação, lançamento, crédito, prescrição e decadência tributários;

Adequado tratamento tributário ao ato cooperativo praticado pelas sociedades cooperativas.

Art. 147. Competem à União, em Território Federal, os impostos estaduais e, se o Território não for dividido em Municípios, cumulativamente, os impostos municipais; ao Distrito Federal cabem os impostos municipais.

Obs: Outros artigos ainda pertencentes a esta Seção I, não fazem falta para a caracterização do contexto, e por isso passa-se agora para a Seção IV.

#### Seção IV

##### Dos Impostos dos Estados e do Distrito Federal

Art. 155. Compete aos Estados e ao Distrito Federal instituir impostos sobre:

- I – transmissão *causa mortis* e doação, de quaisquer bens ou direitos ;
- II – operações relativas à circulação de mercadorias e sobre prestações de serviços de transporte interestadual e intermunicipal e de comunicação, ainda que as operações e as prestações se iniciem no exterior;
- III – propriedade de veículos automotores

&1º. (Não será utilizado neste trabalho)

&2º. O imposto previsto no inciso II – **ICMS**, atenderá ao seguinte:

I – será não cumulativo, compensando-se o que for devido em cada operação relativa à circulação de mercadorias ou prestação de serviços com o montante cobrado nas anteriores pelo mesmo ou outro Estado ou pelo Distrito Federal;

II – a isenção ou não incidência, salvo determinação em contrário da legislação:

Não implicará crédito para compensação com o montante devido nas operações ou prestações seguintes;

Acarretará a anulação do crédito relativo às operações anteriores;

III – Poderá ser seletivo, em função da essencialidade das mercadorias e dos serviços;

IV – resolução do Senado Federal, de iniciativa do Presidente da República ou de um terço dos Senadores, aprovada pela maioria absoluta dos seus membros, estabelecerá as alíquotas aplicáveis às operações e prestações, interestaduais e de exportação.

V – É facultado ao Senado Federal:

Estabelecer alíquotas mínimas nas operações internas, mediante resolução de iniciativa de um terço e aprovada pela maioria absoluta de seus membros.

Estabelecer alíquotas máximas nas mesmas operações para resolver conflito específico que envolva interesse de Estados, mediante resolução de iniciativa da maioria absoluta e aprovada por dois terços de seus membros.

VI – salvo deliberação em contrário, dos Estados e do Distrito Federal, nos termos do disposto no inciso XII, “g”, as alíquotas internas, nas operações relativas à circulação de mercadorias e nas prestações de serviços, não poderão ser inferiores às previstas para as operações interestaduais.

VII – em relação às operações e prestações que destinem bens e serviços a consumidor final localizado em outro Estado, adotar-se-á:

a alíquota interestadual, quando o destinatário for contribuinte do imposto;

a alíquota interna, quando o destinatário não for contribuinte dele;

VIII – na hipótese da alínea “a” do inciso anterior, caberá ao Estado da localização do destinatário, o imposto correspondente à diferença entre a alíquota interna e a interestadual;

IX – incidirá também:

Sobre a entrada de mercadoria importada do exterior, ainda quando se tratar de bem destinado a consumo ou ativo fixo do estabelecimento, assim como sobre serviço prestado no exterior, cabendo o imposto ao Estado onde estiver situado o estabelecimento destinatário da mercadoria ou do serviço;

sobre o valor total da operação, quando mercadorias forem fornecidas com serviços não compreendidos na competência tributária dos Municípios;

X – não incidirá:

Sobre operações que destinem ao exterior produtos industrializados, excluídos os semi elaborados definidos em lei complementar;

Sobre operações que destinem a outros Estados petróleo, inclusive lubrificantes, combustíveis líquidos e gasosos dele derivados, e energia elétrica;

Sobre o ouro, nas hipóteses definidos no art. 153 &5º;

XI – não compreenderá, em sua base de cálculo, o montante dos impostos sobre produtos industrializados, quando a operação, realizada entre contribuintes e relativa a produto destinado a industrialização ou à comercialização, configure fato gerador de dois impostos;

XII – cabe à lei complementar:

Definir seus contribuintes;

Dispor sobre substituição tributária;

Disciplinar o regime de compensação do imposto;

Fixar, para efeito de sua cobrança e definição do estabelecimento responsável, o local das operações relativas à circulação de mercadorias e das prestações de serviços;

Excluir da incidência do imposto, nas exportações para o exterior, serviços e outros produtos além dos mencionados no inciso X, “a”;

Prever os casos de manutenção de crédito, relativamente à remessa para outro Estado e exportação para o exterior, de serviços e mercadorias;

Regular a forma como, mediante deliberação do Estado e do Distrito Federal, isenções, incentivos e benefícios fiscais serão concedidos e revogados;

&3º. a exceção dos impostos de que tratam o inciso II do caput deste artigo e o art 153,

I e II, nenhum outro tributo poderá incidir sobre operações relativas a energia elétrica, serviços de telecomunicações, derivados de petróleo, combustíveis e minerais do País;